

Trabalho Prático

Laboratório de Big Data

Mestrado em Engenharia Informática - Engenharia de Dados
2º ano 1º semestre
2024/2025

1. Objetivos

2. Cronograma

3. Regulamentos

3.1 Artigo Científico

3.2 Avaliação

4. Descrição do Trabalho

5. Referências Bibliográficas

1. Objetivos

Objetivo Geral:

-Propor uma arquitetura para um projeto de Big Data em diferentes domínios de aplicação.

Objetivos Específicos:

- Definir o problema e descrever um projeto de Big Data a analisar
- Faça uma revisão da literatura sobre o problema e das contribuições anteriores para o domínio do projeto de Big Data selecionado
- Selecione tecnologias e estruturas
- Propor um protótipo de arquitetura para a resolução do problema (componentes, tecnologias e frameworks a utilizar em cada componente)
- Definir a metodologia de trabalho
- Escrever um Artigo Técnico

2. Cronograma

Divulgação de propostas de trabalho: até 4 de outubro de 2024
Envio de trabalho: até 27 de outubro de 2024 (23:59)
Defesa e discussão: a ser agendado pelo professor de PL.

3.º Regras

- O grupo (máximo de 3 membros) deve ser o mesmo em ambas as iterações do Trabalho Prático.
- O Python deve ser utilizado como ferramenta de suporte ao processamento de dados.
- O prazo para SUBMISSÃO do Trabalho Prático é 27 de outubro de 2024, no Moodle. Independentemente deste prazo, os grupos deverão poder reportar o estado de desenvolvimento do Trabalho Prático quando solicitado pelo docente.
- O Trabalho Prático é constituído por um Artigo Científico.
- Deve enviar todos os documentos num arquivo compactado. O ficheiro zip deve conter:
 - Ó o artigo científico em conjuntos de dados em
 - Ó formato PDF
 - Ó apresentação em ppt (10 minutos)
- O nome do ficheiro deve seguir a seguinte notação:
LABGDD_YYY_XXX_StudentNo1_StudentNo2_StudentNo3.zip, onde AAA representa a sigla do professor PL, e XXX representa a classe PL.

Exemplo: LABGDD_AMD_M1A_7777777_8888888_9999999.zip.
- Obras cuja designação não respeite a notação indicada será penalizado em 10%.
- Os trabalhos práticos deverão ser entregues no Moodle no prazo definido. Não serão aceites submissões tardias.
- A defesa e discussão dos trabalhos práticos decorrerá em dia e hora a agendar por cada docente das aulas laboratoriais. No dia da apresentação, TODOS os membros do trabalho em equipa devem estar presentes. Os associados ausentes não receberão nota. A defesa e discussão serão conduzidas em equipa com questões dirigidas a cada membro individualmente.
- Cada equipa de trabalho é responsável pela gestão do seu processo de desenvolvimento. As dificuldades e problemas devem ser comunicados ao professor da turma laboratorial em tempo útil.
- Código de conduta: (cf. Regulamento Disciplinar dos Estudantes do IPP)
 - Ó Nenhum aluno ou equipa de trabalho pode reivindicar a propriedade de trabalhos realizados por terceiros ou desenvolvidos em colaboração com terceiros.
 - Ó É expressamente proibida a utilização de materiais, artefactos ou códigos de terceiros sem a devida e explícita indicação de origem.
 - Ó O código de outras fontes deve ser claramente identificado dentro do próprio código, indicando a fonte.
 - Ó Os casos de apropriação não autorizada de materiais, artefactos e/ou códigos sujeitos a avaliação serão comunicados ao Reitor do ISEP.
 - Ó Deve ser mencionada a utilização de ferramentas de auxílio à codificação/desenho de IA (por exemplo, ChatGPT).

-O uso da ferramenta de controlo de versões Bitbucket é obrigatório.

3.1. Artigo Científico

O artigo científico deverá documentar todas as fases da metodologia de trabalho, definição do problema, revisão da literatura, arquitetura proposta e conclusões (máximo de 8 páginas utilizando o template IEEE disponibilizado no Moodle). Considere os seguintes aspetos:

- O artigo deverá ter uma secção inicial com um resumo da motivação e objectivos do trabalho e uma secção final com um resumo das principais conclusões retiradas da resolução das diferentes questões identificadas no enunciado.
- Descrever os componentes para o desenvolvimento da arquitetura Big Data: definição do problema e do projeto Big Data; revisão da literatura; seleção de tecnologias e frameworks; Proposta de arquitetura de Big Data. Uma síntese do trabalho desenvolvido deverá ser incluída na secção de conclusões finais.

3.2. Avaliação

Na avaliação do Trabalho Prático serão considerados os seguintes aspetos e pesos especificados na Tabela 1:

- um)Contextualização e objetivos (Resumo e Introdução)
- b)Revisão da literatura
- c)Significância e complexidade do conjunto de dados
- e)Organização, qualidade da redação, apresentação e clareza do artigo científico
- e)Defesa e discussão
- e)Participação individual de cada elemento do grupo

Tabela 1 – Grelha de Avaliação dos Trabalhos Práticos

Resumo e introdução	15%
Definição do problema	10%
Revisão da literatura	25%
Arquitetura de Big Data	35%
Conclusões e referências	15%

Nota:A nota de cada elemento do grupo será determinada com base na sua participação (em %). A equipa de avaliação dos trabalhos práticos validará, no momento da defesa do trabalho (que poderá ser através de videoconferência), a participação de cada membro do grupo na concretização do trabalho e dos objetivos do grupo.Os associados ausentes não receberão nota.

4. Descrição do Trabalho

Com o desenvolvimento do Trabalho Prático espera-se que os alunos desenvolvam competências no desenvolvimento de abordagens/soluções para projetos de Big Data [1-3], culminando na redação de um artigo científico com a definição do problema, uma revisão de literatura, a arquitetura proposta e as conclusões.

O principal objetivo é fomentar competências no desenvolvimento de arquiteturas de Big Data em diferentes domínios de aplicação.

5. Referências Bibliográficas

[1].Ian Goodfellow, Yoshua Bengio e Aaron Courville, Deep Learning, MIT Press, 2016 (<https://www.deeplearningbook.org/>).

- [2].P. Zikopoulos e C. Eaton, Understanding big data: analysis for business class hadoop and data streaming McGraw-Hill Osborne Media, 2011.
- [3].B. Chambers & M. Zaharia, Spark: O guia definitivo: processamento simplificado de big data, O'Reilly Media, Inc, 2018.- Recursos online e documentação de Tensorflow, Hadoop, Pandas, Numpy, Apache Spark e outros.
- [4].Recursos online e documentação do Tensorflow (<https://www.tensorflow.org/overview>)