**Escola Superior de Tecnologia de Tomar** **Ano letivo:** 2018/2019

**Mestrado em Engenharia Informática - Internet das Coisas**
Mestrado, 2º Ciclo
Plano: Despacho n.º 7043/2016 - 27/05/2016

**Ficha da Unidade Curricular: Análise e Processamento de Grandes Volumes de Dados**
ECTS: 7.5; Horas - Totais: 203.0, Contacto e Tipologia, TP:30.0; PL:30.0; OT:15.0; O:10.0;
Ano | Semestre: 1 | S1
Tipo: Obrigatória; Interação: Presencial; Código: 39091
Área Científica: Ciências e Tecnologias da Programação

**Docente Responsável**
Ricardo Nuno Taborda Campos
Professor Adjunto

**Docente(s)**
Ricardo Nuno Taborda Campos
Professor Adjunto

**Objetivos de Aprendizagem**

1. Conhecer as cinco dimensões do big data
2. Entender os riscos no uso do big data
3. Entender a arquitetura de um sistema de big data
4. Extrair e analisar informação a partir de fontes de dados

**Objetivos de Aprendizagem (detalhado)**

1. Conhecer as cinco dimensões do big data
2. Entender os riscos no uso do big data
3. Entender o ciclo de vida de um projeto de big data bem como a sua arquitetura
4. Entender o processo de query, armazenamento e processamento por detrás do big data
5. Extrair informação a partir de fontes de dados

**Conteúdos Programáticos**

1. Introdução ao Data Science
2. Ética e Privacidade de dados
3. Introdução ao big data
4. Framework de armazenamento e processamento em big data: Apache Hadoop
5. Análise de dados em big data com Python

**Conteúdos Programáticos (detalhado)**

1. Introdução ao Data Science
- O que é o Data Science?
- Data Analysis, Data Analytics, Big Data
- Competências para se tornar um Data Scientist
- Ciclo de vida de um projeto em Data Science

2. Ética e Privacidade de dados
- Como é que podemos evitar o big data?
- Identidade;
- Privacidade;
- Ética;
- Propriedade;
- Reputação;

3. Introdução ao big data
- O que é o big data?
- Quem está a usar Big Data?
- Origens da informação;
- Razões para colecionar tantos dados;
- Como é que o big data difere das tradicionais bases de dados?
- Diferentes tipos de dados;
- 5 V?s do Big Data: volume, velocidade, variedade, veracidade e valor;

4. Framework de armazenamento e processamento em big data: Apache Hadoop e Spark
- HDFS;
- MapReduce;
- RDDs
- Dataframes
- Streaming

5. Análise de dados em big data
- O que é?
- Aplicações;
- Arquitetura de um sistema de processamento da linguagem natural;
- Soluções comerciais;
- Text Analytics com Python

**Metodologias de avaliação**

Avaliação periódica: Trabalho de Investigação (50%) + Hands on Lab (50%)

A entrega do trabalho de investigação e do hands on lab é obrigatória para a obtenção de aprovação na unidade curricular durante a avaliação periódica que pressupõe um mínimo de 70% de presenças. A entrega fora do prazo previsto implica a reprovação automática do aluno impossibilitando-o de se propor a exame. Os alunos ficam também automaticamente reprovados e excluídos de exame no caso de obterem nota inferior a 4 valores em qualquer um dos dois momentos de avaliação (trabalho de investigação ou hands on lab) ou no caso de não atingirem um número mínimo de presenças.

Avaliação Final: Trabalho de Investigação (100%)

os interessados em obter aprovação/melhoria na UC durante a época de exame deverão proceder à melhoria do trabalho de investigação realizado durante a avaliação periódica, cuja pontuação será assim adaptada a 100%

**Software utilizado em aula**

Apache Hadoop; Spark; Python: Anaconda e Jupyter Notebooks

**Estágio**

Não aplicável

**Bibliografia recomendada**

- Provost, F. e Fawcett, T. e , . (2013). *Data Science for Business* (pp. 1-386).USA: OÂ´Reilly
- Witten, I. e Frank, E. e Hall, M. (2011). *Data Mining: Practical Machine Learning Tools and Techniques* (pp. 1-629).USA: Elsevier
- Erl, T. e Khattak, W. e Buhler, P. (2016). *Big Data Fundamentals: Concepts, Drivers & Techniques* (pp. 1-235).USA: Prentice Hall
- Davis, K. (2012). *Ethics of Big Data* (pp. 1-79).USA: OÂ´Reilly

**Coerência dos conteúdos programáticos com os objetivos**

O programa cobre os diferentes objectivos e competências específicas que se pretendem proporcionar na unidade curricular, de acordo com a correspondência seguinte: Objectivos e competências/conteúdos programáticos
Objectivos 1: Conteúdos 1
Objectivos 2: Conteúdos 2
Objectivos 3: Conteúdos 3, 4
Objectivos 4: Conteúdos 5

**Metodologias de ensino**

Ensino teórico-prático com recurso a meios áudio-visuais, a equipamento laboratorial e a

exemplos práticos. Avaliação: Realização e apresentação de projectos de grupo.

**Coerência das metodologias de ensino com os objetivos**

Os objectivos da unidade curricular são atingidos através de um conjunto diversificado de actividades educativas e de avaliação, que preparam e enquadram o trabalho autónomo do estudante pela transmissão de saberes teóricos, técnicos e metodológicos em contexto de aula, de orientação tutorial, de supervisão individual, de participação em grupos, projectos e redes de investigação.

**Língua de ensino**

Português

**Pré-requisitos**

Não aplicável

**Programas Opcionais recomendados**

Não aplicável

---

**Docente responsável**

Ricardo Campos

Assinado de forma digital por Ricardo Campos
DN: cn=Ricardo Campos, o=IPT, ou=ESGT, email=ricardo.campos@ipt.pt, c=PT
Dados: 2019.02.21 15:43:23 Z