



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; Ramo: Tronco comum;

Tipo: Obrigatória; Interação: Presencial; Código: 39091

Área Científica: Ciências e Tecnologias da Programação

Docente Responsável

Ricardo Nuno Taborda Campos

Docente e horas de contacto

Ricardo Nuno Taborda Campos

Professor Adjunto, TP: 30; PL: 30; OT: 15;

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 e Spark
5. Análise de dados em big data

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

5. Análise de dados em big data

- O que é?
- Aplicações;
- Arquitetura de um sistema de processamento da linguagem natural;
- Soluções comerciais;
- Text Analytics: Python e NLTK

Metodologias de avaliação

Avaliação periódica: Projeto (100%)

A entrega do projeto é obrigatória para a obtenção de aprovação na unidade curricular durante a avaliação periódica. No caso de não entrega durante o prazo previsto, os alunos ficam automaticamente reprovados, ficando impossibilitados de se propor a exame. Os alunos ficam também automaticamente reprovados e excluídos de exame no caso de obterem nota inferior a 7 valores.

Avaliação final (época de exames): Projeto (100%)

Os alunos que não obtenham aproveitamento à unidade curricular durante a avaliação periódica ($7 \leq \text{nota} < 9.5$) ou que tendo obtido aproveitamento ($\text{nota} \geq 9.5$) desejem fazer melhoria, deverão submeter-se à avaliação final durante a época de exames.

Software utilizado em aula

Apache Hadoop; Spark; Python

Estágio

Não aplicável

Bibliografia recomendada

- Davis, Kord (2012). Ethics of Big Data. USA: O'Reilly.
- deRoos, Dirk, Paul C. Zikopoulos, Bruce Brown, Rafael Coss, e Roman B. Melnyk (2014). Hadoop for Dummies. USA: Wiley.
- Erl, Thomas, Wajid Khattak, e Paul Buhler (2016). Big Data Fundamentals: Concepts, Drivers & Techniques. USA: Prentice Hall.



- Fowler, Adam (2015). NoSQL For Dummies. USA: Wiley.
- Holmes, Alex. Hadoop in Practice (2012). USA: Manning Publications.
- Hurwitz, Judith, Alan Nugent, Fern Halper, e Marcia Kaufman (2013). Big Data For Dummies. USA: Wiley.
- Marr, Bernard (2015). Using Smart Big Data Analytics and Metrics To Make Better Decisions and Improve Performance. United Kingdom: Wiley.
- Owens, Jonathan R., Jon Lentz, e Brian Femiano (2013). Hadoop Real-World Solutions Cookbook. UK: Packt Publishing.
- Prajapati, Vignesh (2013). Big Data Analytics with R and Hadoop. UK: Packt Publishing.
- Provost, Foster, e Tom Fawcett (2013). Data Science for Business. USA: O'Reilly.
- Sitto, Kevin, e Marshall Presser (2015). Field Guide to Hadoop. USA: O'Reilly.
- Sremack, Joe (2015). Big Data Forensics – Learning Hadoop Investigations. UK: Packt Publishing.
- Stackowiak, Robert, Art Licht, Venu Mantha, e Louis Nagode (2015). Big Data and the Internet of Things. USA: Apress.
- Turkington, Garry (2013). Hadoop Beginner's Guide. UK: Published by Packt Publishing Ltd.
- Witten, Ian H., Eibe Frank, e Mark A. Hall (2011). Data Mining: Practical Machine Learning Tools and Techniques. USA: Elsevier

Coerência dos conteúdos programáticos com os objetivos

O programa cobre os diferentes objetivos e competências específicas que se pretendem proporcionar na unidade curricular, de acordo com a correspondência seguinte: Objetivos e competências/conteúdos programáticos

Objetivos 1: Conteúdos 1

Objetivos 2: Conteúdos 2

Objetivos 3: Conteúdos 3, 4

Objetivos 4: Conteúdos 5

Metodologias de ensino

Ensino teórico-prático com recurso a meios audiovisuais, a equipamento laboratorial e a exemplos práticos.
Avaliação: Realização e apresentação de projeto de grupo.

Coerência das metodologias de ensino com os objetivos

Os objetivos da unidade curricular são atingidos através de um conjunto diversificado de atividades 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, projetos e redes de investigação.

Língua de ensino

Português

Pré requisitos

Não aplicável

Programas Opcionais recomendados

Não aplicável

Observações

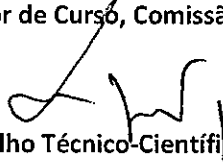
Ricardo
Campos

Assinado de forma digital por Ricardo
Campos
DN: cn=Ricardo Campos, o=IPT,
ou=ESGT, email=ricardo.campos@ipt.pt,
c=PT
Dados: 2017.12.18 18:04:52 Z

Docente Responsável



Diretor de Curso, Comissão de Curso



Conselho Técnico-Científico

10/12/2017 18:04:52 Z

4