



INSTITUTO POLITÉCNICO DE COIMBRA
INSTITUTO SUPERIOR DE CONTABILIDADE E ADMINISTRAÇÃO DE COIMBRA

Aprovação do Conselho Pedagógico

20/3/2019

Aprovação do Conselho
Técnico-Científico

10/4/2019

Ficha de Unidade Curricular (FUC) de Big Data

Curso(s): Mestrado em Análise de Dados e Sistemas de Apoio à Decisão

Ano Curricular: 1ºano

Trimestre curricular: 4º Trimestre

Ano lectivo: 201819

Docente Responsável: Pedro João Coimbra Martins

Unidade Curricular

Designação:	Big Data
Curso(s):	Mestrado em Análise de Dados e Sistemas de Apoio à Decisão
Ano curricular:	1ºano
Semestre curricular:	4º Trimestre
Número de ECTS:	5
Horas de contacto:	30

1. Corpo Docente

***Replicar pelo nº de docentes, sendo que o primeiro deverá ser o responsável da UC.

Nome: Pedro João Coimbra Martins

Gabinete: 2.09

Email: pmartins@iscac.pt

Nome: António Rui Trigo Ribeiro

Gabinete: 2.18

Email: aribeiro@iscac.pt

2.Funcionamento

a) Objectivos:

A geração e armazenamento de dados tem vindo a registar um assinalável crescimento, obrigando a processos de gestão e tratamento da informação cada vez mais complexos e abrangentes. Esta dinâmica observa-se nas mais variadas áreas, desde dados recolhidos nos cartões de fidelização, a dados disponibilizados em redes sociais em ambiente virtual, passando também por dados gerados automaticamente por aparelhos de análise clínica, entre outros. Toda essa informação abre um vasto número de oportunidades, nomeadamente para as empresas, para a clínica, e para o conhecimento em geral. Pretende-se que o aluno conheça as mais usuais ferramentas informáticas para gestão e armazenamento de dados, conheça e saiba aplicar metodologias de optimização em redes para análise e interpretação de dados de grande dimensão.

b) Regime de frequencia e metodologia de avaliação:

A actividade lectiva decorre em regime presencial, com exposição de conceitos, técnicas e métodos, com grande enfoque em aplicações práticas. Será utilizado software adequado a cada um dos temas propostos.

O aluno será avaliado por um trabalho cujo âmbito abrange os dois temas principais do programa. Os trabalhos são realizados individualmente. A classificação final incidirá na avaliação de um relatório escrito, na avaliação dos elementos de suporte desenvolvidos e na apresentação oral do trabalho.

c) Programa:

1 - Gestão de Big Data

1.1 - Introdução, origens e exemplos

1.2 - Bases de dados e álgebra relacional

1.3 - Bases de dados paralelas

1.4 - MapReduce e Hadoop

1.5 - Bases de dados NoSQL

2 - Metodologias de optimização em redes

2.1 - Caracterização de redes/grafos

2.2 - Medidas de distância e semelhança

2.3 - Técnicas de clustering hierárquico

2.4 - Técnicas de clustering aglomerativo

2.5 - Aplicações da optimização em redes à pesquisa em dados de grande dimensão

d) Bibliografia:

Doug Laney, “3D Data Management: Controlling Data Volume, Velocity, and Variety”, Gartner, February 2001.

Stonebraker et al., “MapReduce and Parallel DBMS’s: Friends or Foes?”, Communications of the ACM, January 2010.

Dean and Ghemawat, “MapReduce: A Flexible Data Processing Tool”, Communications of the ACM, January 2010.

Rick Cattell, “Scalable SQL and NoSQL Data Stores”, SIGMOD Record, December 2010 (39:4).

Ahuja, R.K., Magnanti, T.L., Orlin, J.B., 1993. Network flows: Theory, algorithms and applications, Prentice-Hall Inc, New Jersey, USA.

Hair, J.F., Tatham, R.L., Anderson, R.E., Black, W., 1998. Multivariate data analysis, Prentice-Hall Inc, New Jersey, USA.

Martins, P., Ladrón, A., Ramalhinho, H., 2014. Maximum cut-clique problem: ILS heuristics and a data analysis application, International Transactions in Operational Research 22(5), 775-809 (DOI: 10.1111/itor.12120).