

Big Data

Ocorrência: 5º semestre

Carga horária: TP 22h30; PL 22h30; OT 7h30

ECTS: 5,0

Área disciplinar: Informática

Objetivos de aprendizagem (conhecimentos, aptidões e competências a desenvolver pelos estudantes):

Os estudantes que terminam com sucesso esta unidade curricular deverão ser capazes de: - Conhecer um conjunto de tópicos não convencionais que permitam uma gestão escalável dos dados, bem como a utilização de algoritmos paralelos e de modelação estatística, com e sem recurso à cloud; - Ser proficientes num ecossistema de ferramentas e plataformas, que lhes permitam, em face de um problema concreto, determinar a solução a aplicar e os instrumentos a utilizar no armazenamento, exploração e análise de um grande volume de dados.

Conteúdos programáticos:

1. Introdução História e contexto. Panorama da tecnologia Big Data. Ciência dos dados. Pesquisa, indexação e memória
2. Manipulação de dados em larga escala Sistema de Armazenamento em Larga Escala. MapReduce e Hadoop. Relação com bases de dados correntes, streams, algoritmos, extensões e linguagens. Processamento paralelo de consultas e análise computacional de estatísticas. Armazenamento chave-valor; comparativo entre bases de dados SQL e não- SQL
3. Comunicação dos resultados Visualização dos resultados computacionais. Proveniência dos dados, privacidade, ética e governação

4. Tópicos Especiais Análise de grafos: estrutura, travessias, análise computacional, PageRank, consultas recursivas, web semântica, sistemas de advertising e recomendação na internet.

Bibliografia principal:

J. Leskovec, A. Rajaraman, J. Ullman, Mining of Massive Datasets, Cambridge University Press, 2nd Ed., 2014

G. Shroff, The Intelligent Web: Search, smart algorithms, and big data, Oxford University Press, 0199646716, 2014

T. White, Hadoop: The Definitive Guide, Yahoo Press, 3rd edition 2012

S. Perera, Instant MapReduce Patterns, Packt Publishing, 2013