

CURSO	Licenciatura em Informática		
UNIDADE CURRICULAR	Tecnologias da Internet IV	Obrigatória	X
		Opcional	
ÁREA CIENTÍFICA	Engenharia Informática		

Ano: 2º	Semestre: 2º	ECTS: 4	Horas de Contacto teórico práticas: 60h
---------	--------------	---------	---

OBJETIVOS DA APRENDIZAGEM

Para concluir com sucesso esta unidade curricular, os estudantes deverão demonstrar possuir os seguintes conhecimentos e capacidades:

1. Compreender os conceitos fundamentais e os elementos da infraestrutura tecnológica de Big Data e Internet of Things (IoT);
2. Compreender as características dos sistemas de alta disponibilidade;
3. Conhecer os elementos da arquitetura e os diferentes cenários de utilização do Hadoop;
4. Saber instalar e configurar o Hadoop;
5. Compreender o modo de funcionamento do HDFS;
6. Saber utilizar os diferentes mecanismos de importação de dados;
7. Saber instalar, configurar e administrar um cluster YARN;
8. Saber administrar clusters Hadoop;
9. Saber instalar e configurar bases de dados HBase

PROGRAMA

1. Conceitos fundamentais sobre Big Data e Internet of Things (IoT)
2. Infraestrutura tecnológica para Big Data e Internet Of Things.
3. Sistemas em cluster de alta disponibilidade
4. A arquitetura Hadoop
 - 4.1. A arquitetura Hadoop e o BigData
 - 4.2. Cenários de utilização do Hadoop
 - 4.3. Componentes do Hadoop
 - 4.4. O Hadoop EcoSystem
 - 4.5. Instalação e configuração do Hadoop
5. O Hadoop Distributed File System (HDFS)
 - 5.1. Arquitetura do HDFS
 - 5.2. Os HDFS Roles: NameNodes, DataNodes e SecondaryNameNodes
 - 5.3. O processo de leitura e escrita de ficheiros
6. Data Ingestion - Importação de dados
 - 6.1. Data Ingestion com o Apache Flume
 - 6.2. Importação de bases de dados relacionais com o Apache Sqoop
 - 6.3. Data Ingestion com REST Web Services

7. Processamento de dados com MapReduce e YARN
 - 7.1. Conceitos base sobre modelo de processamento de dados MapReduce
 - 7.2. Conceitos base e componentes da arquitetura Hadoop Yarn:
 - 7.3. Instalação e configuração de clusters Hadoop Yarn
8. Administração de um cluster Hadoop
 - 8.1. Administração de clusters YARN e execução de aplicações MapReduce
 - 8.2. Administração do HDFS
 - 8.3. Monitorização do cluster e auditoria
 - 8.4. Implementação de mecanismos de segurança
9. Utilização componentes de processamento análise de dados do Hadoop EcoSystem
 - 9.1. Instalação e configuração do Apache Hive
 - 9.2. Instalação e configuração do Apache Pig
 - 9.3. Instalação e configuração do Apache Spark
10. Bases de dados HBase e Hadoop
 - 10.1. Tipos de bases de dados NoSQL
 - 10.2. Componentes base do HBase
 - 10.3. O modelo de dados do Hbase
 - 10.4. Distribuição de dados HBase e arquitetura de cluster
 - 10.5. Instalação e configuração do HBase
 - 10.6. Integração entre HBase e Hadoop
11. Administração de sistemas Hbase
 - 11.1. Gestão de clusters Hbase
 - 11.2. Monitorização do funcionamento do Hbase
 - 11.3. Técnicas de otimização de bases de dados
 - 11.4. Resolução de problemas

DEMONSTRAÇÃO DE COERÊNCIA ENTRE CONTEÚDOS PROGRAMÁTICOS E RESULTADOS DA APRENDIZAGEM

O objetivo 1 é concretizado através dos pontos 1 e 2. O ponto 3 permite concretizar o objetivo 2. Os objetivos 3 e 4 são concretizados através dos pontos 4.1 a 4.5. Os pontos 5.1 a 5.3 permitem a realização do objetivo 5. O objetivo 6 é concretizado através dos pontos 6.1 a 6.3. Os pontos 7.1 a 7.3 permitem alcançar o objetivo 7. O objetivo 8 é concretizado através dos pontos 8.1 a 8.4 e 9.1 a 9.3. Os pontos 10.1 a 10.6 e 11.1 a 11.4 permitem concretizar o objetivo 9.

METODOLOGIA DE ENSINO E AVALIAÇÃO

Esta unidade curricular tem uma natureza teórico-prática. Estão previstas 60 horas de contato. O tempo total de trabalho do aluno corresponde a 108 horas.

As matérias de enquadramento teórico são complementadas com projetos práticos de implementação da infraestrutura de hardware e software para o processamento de dados do tipo BigData.

De acordo com o Regulamento de Funcionamento do ISTECS a avaliação é efetuada através de um exame escrito individual e obrigatório. Na classificação final, poderão ser considerados

elementos de avaliação contínua, tais como testes, trabalhos individuais ou em grupo, assim como a participação nas aulas presenciais e em recursos de aprendizagem proporcionados por sistemas de e-learning.

DEMONSTRAÇÃO DE COERÊNCIA ENTRE METODOLOGIAS DE ENSINO E RESULTADOS DE APRENDIZAGEM

A componente teórica e a prática laboratorial de implementação de soluções de cluster para processamentos BigData, asseguram que os alunos adquirem os conhecimentos teóricos e as competências práticas definidas como objetivos da unidade curricular.

BIBLIOGRAFIA

Fundamental:

Rajkumar Buyya; Amir Vahid Dastjerdi, Internet of Things, Morgan Kaufmann 2016

Robert Stackowiak; Art Licht; Venu Mantha; Louis Nagode, Big Data and the Internet of Things: Enterprise Information Architecture for a New Age, Apress 2015

Tom White, Hadoop: The Definitive Guide, O'Reilly 2015

Nishant Garg, HBase Essentials, Packt 2014

Complementar:

Sander van Vugt, Pro Linux High Availability Clustering, Apress 2014

Sam R. Alapati, Expert Hadoop Administration: Managing, Tuning, and Securing Spark, YARN, and HDFS, Addison-Wesley Data & Analytics 2016

Gurmukh Singh, Monitoring Hadoop, Packt Publishing 2015

Danil Zburivsky, Hadoop Cluster Deployment, Packt Publish 2013

Ben Spivey; Joey Echeverria, Hadoop Security, O'Reilly 2015