



## Plano de Ensino para o Ano Letivo de 2021

IDENTIFICAÇÃO		
Disciplina: Introdução Big Data		Código da Disciplina: MIN705
Course: -		
Materia:		
Periodicidade: Semestral	Carga horária total: 40	Carga horária semanal: 00 - 00 - 02
Curso/Habilitação/Ênfase:	Série:	Período:
Administração	4	Matutino
Administração	4	Noturno
Engenharia de Alimentos	5	Diurno
Engenharia de Controle e Automação	5	Diurno
Engenharia de Controle e Automação	6	Noturno
Engenharia de Controle e Automação	5	Noturno
Engenharia de Computação	5	Diurno
Engenharia Civil	5	Diurno
Engenharia Civil	6	Noturno
Engenharia Civil	5	Noturno
Design	4	Matutino
Design	4	Noturno
Engenharia Eletrônica	5	Diurno
Engenharia Eletrônica	6	Noturno
Engenharia Elétrica	6	Noturno
Engenharia Elétrica	5	Diurno
Engenharia Mecânica	5	Diurno
Engenharia Mecânica	6	Noturno
Engenharia Mecânica	5	Noturno
Engenharia de Produção	6	Noturno
Engenharia de Produção	5	Diurno
Engenharia de Produção	5	Noturno
Engenharia Química	6	Noturno
Engenharia Química	5	Diurno
Engenharia Química	5	Noturno
Professor Responsável:	Titulação - Graduação	Pós-Graduação
Murilo Zanini de Carvalho	Tecnologia em Eletrônica	Mestre
Professores:	Titulação - Graduação	Pós-Graduação
Murilo Zanini de Carvalho	Tecnologia em Eletrônica	Mestre
Tiago Sanches da Silva	Engenheiro em Elétrica e Eletrônica	Mestre



MODALIDADE DE ENSINO
<p>Presencial: 100%</p> <p>Mediada por tecnologia: 0%</p> <p>* Em qualquer modalidade a entrega de atividades e trabalhos deve ser realizada segundo orientações do professor da disciplina.</p>
ATIVIDADES DE EXTENSÃO
A DISCIPLINA NÃO CONTEMPLA ATIVIDADES DE EXTENSÃO.
EMENTA
<p>Conceitos e definições de Big Data. Principais características de Big Data. Introdução ao Gerenciamento e armazenamento da informação. Características de uma plataforma Big Data. Tecnologias associadas à Plataforma Big Data. Dados Estruturados e Dados Não-Estruturados. Modelos de Serviços em Nuvem. Paralelização de processamento (Map Reduce). Ferramenta Hadoop, Hive, Spark. Introdução ao NoSQL.</p>
SYLLABUS
<p>Big Data concepts and definitions. Main features of Big Data. Introduction to information management and storage. Features of a Big Data platform. Technologies associated with the Big Data Platform. Structured Data and Unstructured Data. Cloud Service Models. Processing parallelization (Map Reduce). Hadoop, Hive, Spark tool. Introduction to NoSQL.</p>
TEMARIO
<p>Conceptos y definiciones de Big Data. Principales características de Big Data. Introducción a la gestión y almacenamiento de información. Características de una plataforma Big Data. Tecnologías asociadas a la Plataforma Big Data. Datos estructurados y datos no estructurados. Modelos de servicios en la nube. Procesamiento de paralelización (Map Reduce). Herramienta Hadoop, Hive, Spark. Introducción a NoSQL.</p>
CONHECIMENTOS PRÉVIOS NECESSÁRIOS PARA O ACOMPANHAMENTO DA DISCIPLINA
<ul style="list-style-type: none"> <li>- Conhecimento intermediário em linguagens de programação e lógica de programação;</li> <li>- Estatística;</li> <li>- Básico de Banco de dados;</li> <li>- Conhecimento básico de protocolos de rede;</li> </ul>
COMPETÊNCIAS DESENVOLVIDAS NA DISCIPLINA
<p><b>COMPETÊNCIA 1:</b></p> <p>Compreender o conceito de DataOps, verificando onde ele deve ser utilizado nos projetos de ciências de dados. Manipular e preparar fluxos de processamento de grandes quantidades de dados.</p>
OBJETIVOS - Conhecimentos, Habilidades, e Atitudes
<ol style="list-style-type: none"> <li>1. Entendimento do conceito de Big Data;</li> <li>2. Visão de todas as áreas que englobam o tema além da computação;</li> <li>3. Plataformas para desenvolvimento do tema;</li> </ol>



4. Entendimento dos diferentes cenários e desafios que os projetos em big data proporcionam;
5. Entendimento sobre dados estruturados e não estruturados, noSQL e técnicas para clusterização.

#### Habilidades

1. Trabalho em equipe;
2. Visão geral do processo de desenvolvimento de uma solução de Big Data;
3. Analisar o contexto do projeto e escolher a melhor solução dentre as inúmeras plataformas disponíveis;
4. Decidir qual o melhor tipo de visualização dos resultados gerados pelo processamento das informações;
5. Criar sistemas de alto desempenho que atendam as restrições impostas por profissionais de ciências de dados.

#### Atitudes

1. Adquirir postura de um profissional ético frente a desafios encontrados nos projetos propostos, enquanto trabalha em equipe;
2. Postura para trabalhar em equipe;
3. Independência e postura de engenheiro em frente a desafios;
4. Interficiar com profissionais de outras áreas;
5. Ter iniciativa para solução de problemas;
6. Criatividade na solução de problemas e desafios.

### ESTRATÉGIAS ATIVAS PARA APRENDIZAGEM - EAA

Aulas de Laboratório - Sim

### LISTA DE ESTRATÉGIAS ATIVAS PARA APRENDIZAGEM

- Peer Instruction (Ensino por pares)
- Project Based Learning
- Problem Based Learning

### METODOLOGIA DIDÁTICA

Os conceitos de Big Data são apresentados aos alunos. O contexto que levou a necessidade de desenvolver tecnologias que trabalham com grande volume de dados é apresentado. Todas as demais aulas utilizam esse contexto apresentado para desenvolver projetos para apresentar os conceitos da disciplina.

Os alunos iniciam a escolha do projeto que eles irão trabalhar ao longo do semestre. A escolha do projeto é viva, recebendo feedback ao longo das entregas realizadas no semestre.

Todos os conceitos e ferramentas para trabalhar com o tópico são apresentados durante as aulas com o intuito de resolver problemas que os alunos já encontraram ou encontrarão nos seus respectivos projetos de Big Data.



INSTRUMENTOS DE AVALIAÇÃO	
NENHUM INSTRUMENTO DE AVALIACAO FOI ADICIONADA.	
AVALIAÇÃO (conforme Resolução RN CEPE 16/2014) e CRITÉRIOS DE APROVAÇÃO	
<p>Disciplina semestral, com trabalhos.</p> <p>Pesos dos trabalhos:</p> <p><math>k_1</math>: 0,4   <math>k_2</math>: 0,6</p>	
INFORMAÇÕES SOBRE INSTRUMENTOS DE AVALIAÇÃO	
CONTRIBUIÇÃO DA DISCIPLINA	
<p>A disciplina exercita conceitos e práticas com o objetivo de identificar oportunidades de exploração de dados muito grandes e complexos (Big Data). Ao seu término o aluno será capaz de planejar e executar iniciativas de Big Data, envolvendo atividades de extração, armazenamento, modelagem e processamento de dados disponíveis na Web e em grandes repositórios.</p>	
BIBLIOGRAFIA	
<p><b>Bibliografia Básica:</b></p> <p>BAESENS, Bart. Analytics in a Big Data World: the essential guide to data science and its applications. Hoboken, N. J: Wiley, c2014. 232 p. ISBN 9781118892701.</p> <p>HURWITZ, Judith et al. Big Data para leigos. Rio de Janeiro: Alta Books, c2016. 301 p. (Tornando tudo mais fácil). ISBN 9788576089551.</p> <p>WHITE, Tom. Hadoop: the definitive guide. 4. ed. Sebastopol, CA: O'Reilly, c2015. 728 p. ISBN 9781491901632.</p> <p><b>Bibliografia Complementar:</b></p> <p>FOWLER, Adam. NoSQL for dummies. Hoboken, N. J: John Wiley &amp; Sons, c2015. 438 p. ISBN 9781118905746.</p> <p>KARANTH, Sandeep. Mastering Hadoop: go beyond the basics and master the next generation of Hadoop data processing platforms. Birmingham, UK: Packt Publishing, c2014. 351 p. ISBN 9781783983643.</p> <p>MARZ, Nathan; WARREN, James. Big Data: principles and best practices of scalable real-time data systems. Shelter Island, NY: Manning, c2015. 308 p. ISBN 9781617290343.</p>	



### SOFTWARES NECESSÁRIOS PARA A DISCIPLINA

- Hadoop Sandbox - Hortonworks Data Platform (HDP) para VMWare (última versão)  
(<https://br.hortonworks.com/products/sandbox/>)
- Docker (última versão)  
(<https://www.docker.com/products/docker-desktop>)
- MongoDB Server (última versão)  
(<https://www.mongodb.com/download-center/community>)
- Cassandra (from datastax) (última versão)  
<https://www.datastax.com/2012/01/getting-started-with-apache-cassandra-on-windows-the-easy-way>
- Anaconda (última versão)  
<https://www.anaconda.com/download/>
- Tensorflow no Anaconda  
<https://anaconda.org/anaconda/tensorflow-gpu>  
<https://www.anaconda.com/blog/developer-blog/tensorflow-in-anaconda/>
- Spark (última versão)  
<https://spark.apache.org/>
- Integrar o Spark com o Python do anaconda usando PySpark  
tuto 1: <https://blog.sicara.com/get-started-pyspark-jupyter-guide-tutorial-ae2fe84f594f>  
tuto 2: <http://media.sundog-soft.com/spark-python-install.pdf>

### INFORMAÇÕES SOBRE PROVAS E TRABALHOS

- k1: Média das avaliação de pesquisas e trabalhos realizados em sala e em casa.
- k2: Projeto final entregue e apresentado no último dia de aula da disciplina.



OUTRAS INFORMAÇÕES



## APROVAÇÕES

Prof.(a) Murilo Zanini de Carvalho

Responsável pela Disciplina

Prof.(a) Angelo Sebastiao Zanini

Coordenador do Curso de Engenharia de Computação

Prof.(a) Cassia Silveira de Assis

Coordenador(a) do Curso de Engenharia Civil

Prof.(a) Claudia Alquezar Facca

Coordenador(a) do Curso de Design

Prof.(a) David Garcia Penof

Coordenador do Curso de Engenharia de Produção

Prof.(a) Edval Delbone

Coordenador(a) do Curso de Engenharia Elétrica

Prof.(a) Eliana Paula Ribeiro

Coordenador(a) do Curso de Engenharia de Alimentos

Prof.(a) Fernando Silveira Madani

Coordenador(a) do Curso de Eng. de Controle e Automação

Prof.(a) Luciano Gonçalves Ribeiro

Coordenador(a) do Curso de Engenharia Química

Prof.(a) Ricardo Balistiero

Coordenador(a) do Curso de Administração

Prof.(a) Sergio Ribeiro Augusto

Coordenador do Curso de Engenharia Eletrônica



Prof.(a) Susana Marraccini Giampietri Lebrao  
Coordenadora do Curso de Engenharia Mecânica

Núcleo Docente Estruturante (NDE)

Data de Aprovação:





## PROGRAMA DA DISCIPLINA

Nº da semana	Conteúdo	EAA
22 L	- PSub 1	0
23 L	- PSub 1	0
24 L	Visão geral do curso e apresentação do plano de ensino	0
25 L	Introdução a Big Data.	61% a 90%
26 L	Gerenciamento e armazenamento da informação. Conceitos gerais de extração de dados.	0
27 L	Introdução ao HADOOP e seu ecossistema.	61% a 90%
28 L	HDFS	11% a 40%
29 L	HiveQL	61% a 90%
30 L	- P3	0
31 L	Trabalho T1 - HiveQL	91% a 100%
32 L	Introdução ao Spark	11% a 40%
33 L	Paralelismo nativo com Spark	41% a 60%
34 L	Projeto de Big Data com Spark	91% a 100%
35 L	Projeto de Big Data com Spark	91% a 100%
36 L	- Eureka	0
37 L	Introdução a NoSQL	0
38 L	- P4	0
39 L	- P4	0
40 L	Apresentação do projeto final de Big Data	0
41 L	- PSub 2	0
Legenda: T = Teoria, E = Exercício, L = Laboratório		