



## MBA em Ciência de Dados (Big Data Processing and Analytics)

### Estrutura Curricular – componente curricular/carga horária

<b>Módulo 1 – Gestão de Dados e Arquitetura de <i>Big Data</i></b>	
Gestão Estratégica de Negócios	32 horas-aula
Cientista de Dados nas Organizações	32 horas-aula
Paradigmas de Linguagens de Programação para Ciência de Dados	32 horas-aula
Arquitetura de <i>BIG DATA</i>	32 horas-aula
<b>Carga horária total do módulo</b>	<b>128 horas-aulas</b>
<b>Módulo 2 – Processamento de <i>Big Data</i></b>	
Coleta e Armazenamento de Dados	32 horas-aula
Recuperação da Informação na Web e em Rede Sociais	32 horas-aula
Visualização de Dados para Tomada de Decisão	32 horas-aula
Processamento em <i>BIG DATA</i> – o <i>Processing</i> em uma Visão Prática	32 horas-aula
<b>Carga horária total do módulo</b>	<b>128 horas-aulas</b>
<b>Módulo 3 – <i>Big Data Analytics</i></b>	
<i>BIG DATA Analytics</i> : Análise Estatística	64 horas-aula
<i>BIG DATA Analytics</i> : Mineração e Análises de Dados	64 horas-aula
<b>Carga horária total do módulo</b>	<b>128 horas-aulas</b>
<b>Módulo 4 – Aplicação de Conhecimento</b>	
Aplicação do Conhecimento	48 horas-aula
<b>Carga horária total do módulo</b>	<b>48 horas-aulas</b>
<b>Total da carga horária do curso</b>	<b>432 horas-aulas</b>



## DISCIPLINA 01

Nome da Disciplina: **Gestão Estratégica de Negócios**

Carga Horária: 32h/aula.

Ementa: Transformação Digital no contexto da Gestão Estratégica de Negócios. Tecnologias Analíticas e Cenários Exponenciais. Arquitetura Corporativa como ferramenta para mapeamento de capabilities e alinhamento estratégico considerando fatores intra e extra organização.

Objetivo: Entender as relações e a importância das ferramentas analíticas no processo de Transformação Digital das organizações para a efetivação de estratégias de negócios, considerando os fatores exponenciais na geração de diferencial competitivo ou disruptivo.

Conteúdo Programático: A Transformação Digital no ambiente de negócios. Missão, Visão organizacional e o PTM (Propósito Transformador Massivo). Criando Estratégias de Negócios com base analítica. Estratégias orientadas a dados. Inovação como ferramenta para disrupção de mercado. Arquitetura Corporativa no mapeamento das *Capabilities* Organizacionais. Análise de Mercado e Competidores. *Cases* de Sucesso.



## DISCIPLINA 02

Nome da Disciplina: **Cientista de Dados nas Organizações**

Carga Horária: 32 h/a

Ementa: Competências e Habilidades do Cientista de Dados. Papéis desempenhados por profissionais de dados nas organizações. A carreira de *Chief Data Officer* e a sua importância no Planejamento Estratégico.

Objetivo: Conhecer os diversos papéis desempenhados por profissionais de Ciência de Dados nas organizações. Distinguir a posição do CDO (*Chief Data Officer*) e a sua participação na definição do planejamento estratégico. Entender a importância de *softskills* como liderança, trabalho em equipe, empatia, comunicação e ética para participar de equipes multidisciplinares e interagir com *stakeholders* de diversas áreas do conhecimento humano.

Conteúdo Programático: Competências e Habilidades do Cientista de Dados. Taxonomia de papéis nas organizações. O CDO como *C-level*, *Chief Data* ou *Chief Digital*? Alinhamento ao Planejamento Estratégico. Liderança e Gestão como competências essenciais. Trabalho multidisciplinar: Equipes, Times e *Squads*. Importância da Comunicação com diversos *stakeholders*. Postura ética e conduta profissional do CDO.



## DISCIPLINA 03

Nome da Disciplina: **Paradigmas de Linguagens de Programação para Ciência de Dados**

Carga Horária: 32 h/a

Ementa: Aspectos básicos de programação das linguagens Python e R para Ciência de Dados. Programação de aplicações, apresentação e desenvolvimento de casos.

Objetivo: Capacitar o aluno a ler, compreender e adapta scripts escritos nas linguagens Python e R e suas principais bibliotecas. Fazer o aluno travar contato com as principais bibliotecas de Python e R para Ciência de Dados e suas aplicações, bem como seus ambientes de desenvolvimento. Capacitar o aluno a desenvolver scripts simples usando essas linguagens e bibliotecas.

Conteúdo Programático: Ambientes de desenvolvimento RStudio e RStudio.cloud. Introdução à programação em R. *Dataframes* em R, seleção e manipulação de dados. Principais *packages* de R para Ciência de Dados. *Tidyr* e *Dplyr*. Seleção e manipulação de dados, *reshape* de dados, implementação de *pipelines*. Visualização de dados com *ggplot2*. Gráficos de Linha e de Distribuição, Linhas de Tendências, Gráficos de Barras, Histogramas e Mapas de Calor. Jupyter Notebook. Introdução à programação em Python. Estruturas de dados básicas, listas, dicionários, funções e arquivos. Principais bibliotecas de Python para Ciência de Dados. *NumPy*. Operações vetoriais e de Álgebra Linear, *reshape* de dados. *pandas*. *Dataframe*, aquisição seleção e manipulação de dados, *merge* e *join*. *Matplotlib* e *seaborn*. Gráficos de Linha e de Distribuição, Gráficos de Barra, Histograma, Gráfico de Densidade, *Boxplot*, Linhas de Tendências e Mapas de Calor. Apresentação de casos selecionados (em Python e/ou R), desenvolvidos pelos alunos e comentados pelo professor.



## DISCIPLINA 04

Nome da Disciplina: **Arquitetura de *BIG DATA***

Carga Horária: 32 h/a

Ementa: Arquitetura de Computação em Nuvem, Big Data na Nuvem (soluções disponíveis), Processamento Massivamente Paralelo (MPP), Segurança da Informação em *Big Data*, Infraestrutura para *Big Data*.

Objetivo: Apresentar os principais componentes da Arquitetura de *Big Data*, as soluções de Computação em Nuvem para *Big Data*, arquitetura e sistemas de processamento massivamente paralelo, e as implicações de segurança envolvendo *Big Data*.

Conteúdo Programático: Computação em Nuvem: Princípios, Modelos de Serviço e Modelos de Implantação. Soluções de Computação em Nuvem para *Big Data*: Soluções de Infraestrutura, Soluções de Plataforma e Soluções de Serviço. Processamento Massivamente Paralelo: Computação de Alto Desempenho, Computação Paralela em Memória Compartilhada e Computação Distribuída. Paradigmas de aplicações em *Big Data*. Segurança da Informação em *Big Data*: Princípios de Segurança da Informação, Principais Desafios de segurança para *Big Data*. Infraestrutura para *Big Data*: Infraestrutura de Hardware, Infraestrutura de Software e Infraestrutura de Recursos Humanos.



## DISCIPLINA 05

Nome da Disciplina: **Coleta e Armazenamento de Dados**

Carga Horária: 32 h/a

Ementa: Processo de aquisição de dados. Modalidades de coleta: manual, automatizada, em lote, em fluxo. Coleta por meio de sensores, Coleta por meio de conectores. Pré-processamento de dados. Persistência em Cluster. Teorema CAP. Paradigmas de persistência NoSQL: Chave-Valor, Orientado a Documentos, Família de Colunas e Grafos. Tendências em coleta e atuação: Internet das coisas, Indústria 4.0 e *Edge Computing*.

Objetivo: Capacitar o aluno para a prática de coleta de dados em lote ou em fluxo, considerando diferentes modalidades de aquisição e ingestão de *raw data*, por meio de sensores ou conectores, de dados ambientais ou de redes sociais. Conhecer o armazenamento em *cluster* e os paradigmas de persistência que apoiam aplicações de *Big Data*, em termos de volume, velocidade e variedade.

Conteúdo Programático: Processo de coleta e aquisição de dados – da ciência para os negócios. Modalidades e mecanismos de coleta de coleta. Pré-processamento. Armazenamento em *Cluster*. Paradigmas de persistência. Tendências em coleta e atuação. *Cases* e exemplos.



## DISCIPLINA 06

Nome da Disciplina: **Recuperação da Informação na Web e em Redes Sociais**

Carga Horária: 32h/aula.

Ementa: Algoritmos e soluções para busca e extração de informação da Web. *Web crawling*. Representações de dados textuais. *Text Mining*. Algoritmos e soluções para a análise de redes sociais *online* e em sites de conteúdo.

Objetivo: Apresentar ao aluno as principais técnicas de Recuperação da Informação na Web. Capacitar o aluno a processar linguagem natural, de modo a permitir-lhe extrair e representar a informação necessária para analisar dados textuais e desenvolver aplicações como análise de sentimento. Apresentar ao aluno os principais conceitos e as principais métricas de Redes Complexas. Capacitar o aluno a usar software para modelagem e análise de redes complexas, permitindo assim que ele desenvolva aplicações em redes sociais.

Conteúdo Programático: Conteúdo Programático: *Web Scraping* com Python. Manipulação de textos em Python, JSON, introdução a Beautiful Soap e ao processamento de linguagem natural com NLTK. Representação de dados textuais. Modelos *bag of words*, *tf-idf* e *word embedding*. Introdução à Análise de Sentimento. Estudo de caso com mensagens do Twitter. Estudos de Caso de *Web Scraping* e *Text Mining*. Introdução à Ciência de Redes Complexas. Motivação e aplicações, tipos de redes e representações de redes. Fundamentos de redes complexas, caminhos, centralidade, agrupamento, *hubs* e *page rank*. Visualização de redes complexas. Análise de Redes Complexas com Networkx. Estudos de Caso de Redes Sociais. Apresentação de casos selecionados e de casos desenvolvidos pelos alunos, comentados pelo professor.



## **DISCIPLINA 07**

Nome da Disciplina: **Visualização de Dados para Tomada de Decisão**

Carga Horária: 32 h/a

Ementa: Exploração de diversas formas de visualização e estudo de como os humanos percebem informação. Visualização de dados numéricos; visualização de dados não numéricos. O processo de criação e interpretação de visualização de dados. Aplicação de técnicas de projeto de interface com usuários na criação de sistemas de visualização.

Objetivo: Apresentar ao aluno formas diferentes e efetivas de representar dados numéricos e não numéricos. Dar ao aluno condições de escolher a forma mais apropriada para apresentar informações de forma visual, de forma a facilitar a obtenção de *insights* e a tomada de decisão. Apresentar ao aluno ferramentas que facilitem o processo de visualização de dados.

Conteúdo Programático: Gráficos em 2D e em 3D. Fotorrealismo. O ser humano: memória, a retina, raciocínio. Percepção em 2D e em perspectiva. Dados, mapeamento e cartas. Coordenadas paralelas e empilhamento de gráficos. Usando cores e as regras de Tufte. Gráficos e redes e suas visualizações. Sistemas geolocalizados. Árvores de mapas e escalas multidimensionais. Sistemas de visualização. Visualização de informação e de bases de dados. Projeto de sistemas de visualização.





## DISCIPLINA 08

Nome da Disciplina: **Processamento em *BIG DATA* – o *Processing* em uma Visão Prática**

Carga Horária: 32h/aula.

Ementa: Processamento em Big Data. Ambientes distribuídos e centralizados. Desafios de banda de rede em ambientes centralizados. Arquiteturas de processamento. Processamento distribuído. Processamento de dados heterogêneos.

Objetivo: Proporcionar aos participantes detalhamento do processamento de dados em ambientes *Big Data*, abordando arquiteturas e características. Explorar de forma aprofundada a utilização de recursos *Big Data* na nuvem. Capacitar os envolvidos a atribuírem arquiteturas de processamento para *Big Data*.

Conteúdo Programático: Ambientes Distribuídos, incluindo *cloud*, ampliação da disponibilidade. Escalabilidade. Ingestão de Dados para *Big Data*. Integração de Dados Estruturados (Extração, Enriquecimento, Transformação e Carga). Construção de Aplicações ETL (*extract/transform/load*). *Data Compression*. Garantia da Qualidade de Dados.



## DISCIPLINA 09

Nome da Disciplina: ***BIG DATA Analytics: Análise Estatística***

Carga Horária: 64 h/aula.

Ementa: Processos de preparação, coleta e tratamento de dados com uso de técnicas e métodos estatísticos específicos. Verificação de aderência a modelos. Estimação. Testes de hipóteses e análises preditivas.

Objetivo: Estudar os conceitos e práticas da análise estatística de dados experimentais, apresentando ferramentas e técnicas que possibilitem a identificação ou predição de fenômenos quantificáveis.

Conteúdo Programático: Preparação de Dados (escala de medidas, variáveis). Coleta de Dados (técnicas de levantamento e amostragem). Análise Exploratória de Dados (tipos de tabulação e representação gráfica). Estatística Paramétrica vs. Não Paramétrica. Estimação de Parâmetros (medidas de tendência central e de dispersão). Distribuições de Probabilidade: Normal, Normal Padrão, t de Student e Quiquadrado. Teste de Hipóteses (testes paramétricos e não paramétricos). Análise de Variância. Análise de Regressão: Simples e Múltipla. Análise Preditiva: análise de categorias (saída discreta/categórica) e previsão de séries temporais (saída contínua).



## DISCIPLINA 10

Nome da Disciplina: ***BIG DATA Analytics: Mineração e Análises de Dados***

Carga Horária: 64h/aula.

Ementa: Conceito de Aprendizagem de Máquina. Conceitos de conjunto de treinamento, generalização, overfitting e validação. Tipos de aprendizagem. Principais tarefas de Aprendizagem de Máquina. Exemplos de Aplicações. O ecossistema Python para Aprendizagem de Máquina. Algoritmos fundamentais.

Objetivo: Entender os conceitos de Aprendizagem de Máquina, generalização, e os métodos para obter generalização. Entender os principais tipos de aprendizagem. Conhecer as principais tarefas que podem ser realizadas com Aprendizagem de Máquina. Travar contato com uma série de aplicações paradigmáticas de Aprendizagem de Máquina. Conhecer e saber utilizar os principais algoritmos de Aprendizagem de Máquina. Compreender o conceito de *Deep Learning*.

Conteúdo Programático: O paradigma de Aprendizagem de Máquina. Aprendizagem não supervisionada, supervisionada e por reforço. Principais tarefas: associação, redução de dimensão, agrupamento, classificação, regressão, detecção de anomalia e previsão de séries temporais. Conjuntos de treinamento, validação e teste. Conceitos de *bias* e variância. *Underfitting*, generalização, *overfitting*, validação e regularização. Introdução aos principais algoritmos de Aprendizagem de Máquina: a priori, análise de componentes principais, k-médias, regressão linear, regressão logística, árvore de decisão, *random forest* e SVM. Introdução a *Deep Learning*. O ecossistema Python para Aprendizagem de Máquina: scikit-learn. Keras.



## DISCIPLINA 11

Nome do Componente Curricular: **Aplicação de Conhecimento**

Carga Horária: 48 horas/aula na modalidade EAD

Ementa: A disciplina promove o desenvolvimento do Trabalho de Aplicação de Conhecimento, com base no método prático e aplicado, o qual direciona o aluno para a resolução de um desafio ou problema real vivenciado em um contexto institucional/pessoal, utilizando os conceitos e práticas abordados ao longo do curso.

Objetivo: Capacitar o participante para investigar, analisar e compreender as causas e as implicações dos desafios em um contexto institucional/pessoal; e com base no diagnóstico e na pesquisa bibliográfica, propor soluções e ações detalhadas, visando à resolução de problemas ou oportunidades reais e pontuais enfrentadas nesse contexto institucional/pessoal.

Conteúdo Programático: Definição do problema/oportunidade/desafio a ser resolvido; Descrição das características gerais do contexto institucional/pessoal; Diagnóstico das origens e implicações do desafio a ser resolvido; Pesquisa bibliográfica sobre os temas relacionados com o desafio do contexto institucional/pessoal; Proposição de soluções e ações detalhadas para a resolução do desafio.