

CIÊNCIA DE DADOS E BIG DATA

Integrado
^ PÓS



Duração

9 meses



Carga horária

360 horas



Metodologia

On-line

SOBRE O CURSO

O Curso de Pós-Graduação em Ciência de Dados e Big Data do Centro Universitário Integrado busca capacitar profissionais por meio da abordagem de processos, técnicas e metodologias atuais, utilizadas na análise de grandes volumes de dados, auxiliando no aprimoramento de habilidades e estratégias para uma atuação no mercado com eficiência e qualidade.

METODOLOGIA

Disciplinas 100% on-line, organizadas em Unidades de Aprendizagem, que oferecem diversificados recursos didáticos, para a abordagem teórica, metodológica e prática do conteúdo.

Conteúdo planejado com base em metodologias ativas para o ensino e aprendizagem, as quais promovem a autonomia do estudante, figura central no processo de construção do conhecimento, bem como a excelência em sua formação.

Acompanhamento personalizado, para orientações, esclarecimento de dúvidas ou solução de eventuais problemas, permitindo interação simples e eficiente com os tutores e demais estudantes do curso, por meio dos fóruns, grupos de WhatsApp, e-mail ou do Serviço de Atendimento no Portal Acadêmico.



DISCIPLINAS E CONTEÚDOS PROGRAMÁTICOS

FUNDAMENTOS DE BIG DATA

- Características de cenários de Big Data
- Hadoop: ecossistema de processamento em Big Data
- Hadoop Distributed Filesystem
- Hadoop Distributed Filesystem na Prática
- Paradigma de programação do MapReduce
- MapReduce no Hadoop: arquitetura de análise distribuída
- Casos de uso do Hadoop no MapReduce
- Apache YARN: gerenciador de recursos no Hadoop MapReduce
- MapReduce na prática: agregação parcial por meio do Combiner para otimizar desempenho
- HIVE: criação de tabelas
- HIVE: consultas por meio da linguagem HIVEQL
- HIVE na prática: uso e comandos no Hadoop



Carga horária: 40 horas

CIÊNCIA DE DADOS

- Introdução à análise de dados
- Inferência estatística
- Aplicações de regressão linear simples
- Correlação com a linguagem R
- Classificação de dados com a linguagem R
- Ciclo de vida de um processo de ciência de dados
- Estatística descritiva para ciência de dados
- Técnicas utilizadas em ciência de dados e Big Data
- Ferramentas utilizadas em ciência de dados e Big Data
- O ecossistema de Big Data
- Papéis dos envolvidos em projetos de Big Data e ciência dos dados
- Aplicação de Big Data



Carga horária: 40 horas



ANÁLISE DE ALGORITMOS

- Modelos de computação e eficiência de algoritmos
- Teorema mestre
- Introdução à análise assintótica
- Problema do segmento de soma máxima
- Problema da mochila
- Problema da maioria
- Problema do escalonamento de intervalos
- Projetos de algoritmos
- Computabilidade
- Análise de algoritmos em problemas P
- Análise de algoritmos em problemas NP-completos
- Análise de algoritmos e ciência dos dados



Carga horária: 40 horas

PREPARAÇÃO E ANÁLISE EXPLORATÓRIA DE DADOS

- Tratando outliers em Pandas e Numpy
- Análise exploratória de dados
- Estatística descritiva usando Pandas e Numpy
- Análise univariada de dados
- Histogramas com Seaborn
- Box-plots com Seaborn
- Análise multivariada de dados
- Gráficos de dispersão com Seaborn
- Gráficos de barras com Seaborn
- Correlações usando Pandas, Numpy e Seaborn
- Tratamento de valores faltantes em Pandas
- Normalização de dados usando Pandas, Numpy e Scikit-learn



Carga horária: 40 horas



BIG DATA E INTERNET DAS COISAS

- Integrando Big data e IOT
- Introdução à ciência de dados
- Ciência de dados e Big Data
- Aplicações da ciência de Dados
- Aprendizado de Máquina (Machine Learning)
- Introdução a Hadoop
- Mineração de textos (Text mining)
- Descoberta de conhecimento com BIG DATA
- Business Intelligence (BI)
- Análise de Dados Utilizando Dashboards
- Inovação e Design Thinking
- Inteligência Artificial



Carga horária: 40 horas

FRAMEWORK DE BIG DATA

- Sharding
- Arquiteturas de hardware/software de big data
- Frameworks que utilizam sharding como forma de distribuição
- Frameworks de Big Data: uma visão geral
- Aplicações simples utilizando frameworks de big data
- Stream de dados: definição e fontes
- Overview de frameworks de stream de big data
- HIVE: arquitetura e componentes
- Frameworks de armazenamento de dados de big data
- Frameworks de armazenamento semiestruturados
- Frameworks de armazenamento não estruturados
- Frameworks de visualização de dados de big data



Carga horária: 40 horas



PROCESSAMENTO PARALELO E DISTRIBUÍDO

- Introdução à programação paralela e distribuída
- Arquiteturas de computadores
- Abordagem para o paralelismo
- Introdução à programação CUDA
- Memória e movimentação de dados em CUDA
- Comunicação entre threads e sincronização em CUDA
- Soma de prefixos (scan)
- Programação distribuída
- Algoritmos paralelos com Rapids
- Algoritmos paralelos com PyCUDA
- Implementação de algoritmos paralelos com Numba Python
- Operações e métodos de listas (Python)



Carga horária: 40 horas

DATA MINING E GRAPH MINING

- Aplicações de data mining
- Agrupamento de dados
- Introdução a data mining
- Classificação de dados
- Aplicação da técnica de classificação de dados
- Regras de associação
- Aplicação da técnicas de associação
- Aplicação da técnicas de agrupamento de dados
- Graph mining
- Mineração de subgrafos
- Classificação de grafos
- Clustering de grafos



Carga horária: 40 horas



PROGRAMAÇÃO EM BIG DATA COM R

- Linguagens multiparadigma: R
- Tipos de dados e variáveis em R
- Atribuição, operadores aritméticos e lógicos em R
- Estrutura condicional em R
- Comando while, variáveis contadoras e acumuladoras em R
- Comando for, variáveis contadoras e acumuladoras em R
- Aplicações utilizando comandos iterativos aninhados em R
- Importação de dados em R
- Visualização de dados em R
- Criação, acesso e operadores de listas em R
- Criação, acesso e operadores de vetores em R
- Operações e métodos de matrizes em R



Carga horária: 40 horas



COMO É A AVALIAÇÃO

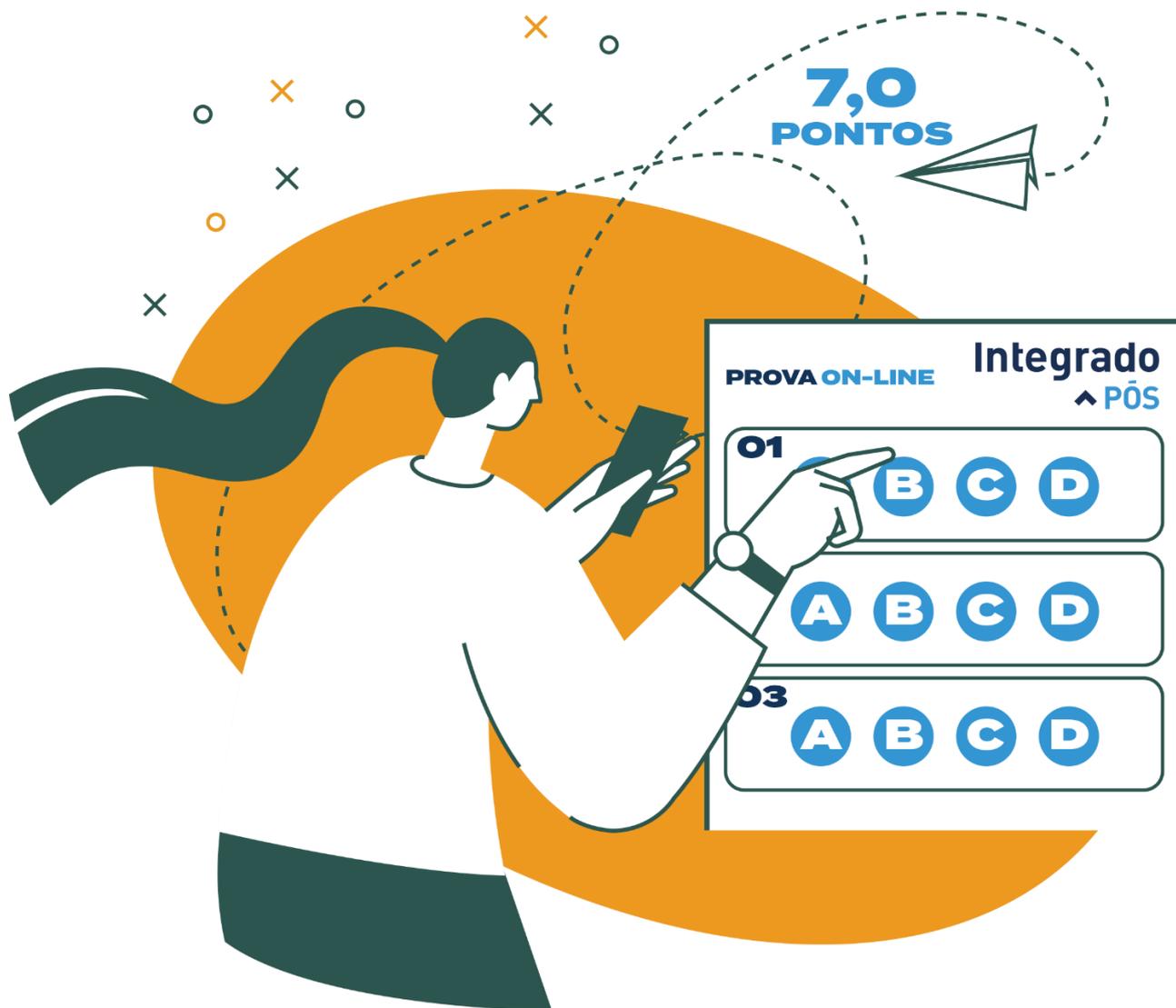
As disciplinas são formadas por Unidades de Aprendizagem, sendo que o seu acesso aos conteúdos disponibilizados em cada uma delas será contabilizado como uma pontuação de progresso, que pode garantir até 3,0 pontos em sua média final. Para isso é necessário:

- ◆ Estudar todos os conteúdos disponibilizados, ou seja, não deixar de acessar nenhuma das *abinhas* que compõem as Unidades de Aprendizagem;
- ◆ Responder corretamente os exercícios propostos.

PROGRESSO NAS UNIDADES DE APRENDIZAGEM



A pontuação de progresso, será somada à nota que você obtiver na prova on-line, formada por questões objetivas de múltipla escolha acerca dos conhecimentos abordados na disciplina, com valor 7,0.



Ao finalizar todas as disciplinas, você terá concluído o curso e seu certificado será emitido em no máximo três meses.

Não é necessário produzir o TCC, mas se você preferir desenvolver o trabalho, lhe ofereceremos todo o suporte necessário.

NÃO PRECISA PRODUZIR O TCC



- ◆ Que tal ser o protagonista do seu próprio conhecimento e se especializar na sua área?
- ◆ Matricule-se já e viva o seu sonho!

