

EXTENSÃO

ENGENHARIA DE

DADOS

48 HORAS

lab.data



FIA - Fundação Instituto de Administração

A FIA dedica-se à geração de conhecimento aplicado e de novos conceitos para a excelência na administração empresarial e pública.

É uma referência de ensino desde o curso de graduação em Administração e Negócios até os cursos de Educação Executiva e em EAD, bem como nos projetos de consultoria e pesquisa aplicada.

A FIA olha para o futuro com determinação, mirando a construção de uma sociedade mais evoluída, consciente e justa, entendendo que o mundo está em evolução, em grande mudança, gerando oportunidades e desafios que construirão uma sociedade coerente com os novos tempos.



BUSINESS SCHOOL

Graduação, Pós-graduação, MBA, Pós-MBA, Mestrado, Profissional, Cursos In-Company e EAD



CONSULTING

Consultoria personalizada que oferece soluções baseadas em seu problema de negócio



RESEARCH

Atualização dos conhecimentos e do material didático oferecidos nas atividades de ensino



Único curso de graduação a receber notas máximas



A primeira escola brasileira a ser finalista da maior competição de MBA do mundo



Única Business School brasileira a figurar no ranking LATAM



Signatária do pacto global da ONU



Membro fundador da ANAMBA – Associação Nacional dos MBAs



Credenciada pela AMBA – Association of MBAs



Credenciada ao Executive MBA Council



Filiada a AACSB – Association to Advance Collegiate Schools of Business



Filiada e EFMD – European Foundation for Management Development



Referência em cursos de MBA nas principais mídias de circulação



3 vezes eleita a melhor escola de Negócios do Brasil pela ÉPOCA



Mais de 8 mil consultorias realizadas em mais de 1 mil empresas



Mais de 40 anos de Experiência com Pesquisa no Brasil e no mundo

Labdata – Laboratório de Análise de Dados

O **Laboratório de Análise de Dados – Labdata** é um Centro de Excelência que atua nas áreas de ensino, pesquisa e consultoria em análise de informação utilizando técnicas de Big Data, Analytics e Inteligência Artificial. É dirigido pela **Prof^a Dra. Alessandra de Ávila Montini**.

Veja porque escolher o Labdata:



Centro de Excelência e pioneiro no lançamento dos cursos de Big Data, Analytics e Inteligência Artificial no Brasil



+ 10 anos de atuação
+9000 alunos formados
+85 turmas de Pós e MBA



Lifelong Learning
Atividades Gratuitas:
Hackathons, Palestras, Cursos de Extensão e Eventos



Número limitado de alunos por turma. **Networking profissional** qualificado e **painel de vagas** com as principais empresas



Cursos de Pós e MBA com carga horária mais completa do mercado, **plantão de dúvidas semanal** com o time de professores do Labdata.



Conteúdo programático completo com **aulas práticas e resolução de cases reais**.



Docentes com sólida formação acadêmica: **doutores e mestres** em sua maioria com larga experiência de mercado



Certificados reconhecidos pelo MEC e **mesma certificação** para os cursos online e presencial



Localização de **fácil acesso** de carro ou transporte público.



Prof^a Dr^a Alessandra Montini

Diretora do Labdata FIA

NOSSOS LABORATÓRIOS

A parceria Labdata e GoWork traz para o ecossistema do coworking, um núcleo de educação executiva, onde pesquisadores, empreendedores e estudantes contribuem para a maximização dos resultados criando um ambiente de busca por conhecimento, sintetizado em muitos eventos, hackathons, aulas de alto nível e amplo espaço para networking. **Localizado na Avenida Paulista**, nossos laboratórios contam com equipamentos de última geração, conforto e segurança para nossos alunos e professores.



Área para Networking



Espaços para Estudos



Salas de Reunião



Áreas de Descompressão



Espaço para Refeições



Quadra de Basquete e Praia



SOBRE O CURSO

OBJETIVO

Após realizar esse curso, os alunos serão capazes de:

- Descrever os mecanismos fundamentais do Apache Spark
- Utilizar as principais APIs do Spark para realizar operações com os dados
- Aprender e utilizar a linguagem de programação Python para criar as aplicações com o Spark
- Implementar casos de uso típicos para o Apache Spark.
- Construir pipelines dos dados e realizar consultas em grandes volumes de dados utilizando o Spark SQL e DataFrames
- Analisar os trabalhos executados pelo Spark através da interface gráfica administrativa, bem como os logs
- Entender o funcionamento interno do Spark
- Processar fluxos de dados em tempo real com escalabilidade, alta vazão e tolerante a falha através do Spark Streaming
- Conhecer as funcionalidades das bibliotecas de aprendizagem de máquina disponível no Spark.

PERFIL DO ALUNO

Profissionais de diversas áreas que desejam entender o framework Apache Spark.

METODOLOGIA

Os conceitos são apresentados por meio de exemplos e exercícios práticos monitorados pelo Professor.



CORPO DOCENTE

O corpo docente conta com professores altamente capacitados com experiência no mundo corporativo. Nos critérios de seleção do corpo docente, serão priorizadas sua qualificação e experiências profissionais nas distintas matérias, de maneira que o curso permita não somente a transmissão de conhecimentos, mas também experiências enriquecedoras para os alunos.

REQUISITOS

- As aulas serão transmitidas através de uma plataforma Digital
- Conexão com a internet - banda larga com ou sem fio com pelo menos 1MB de velocidade final, e para acompanhar as aulas ao vivo e para a realização de exercícios
- Computador com configuração mínima: Dual Core 2Ghz ou superior (i3/i5/i7 ou AMD equivalente) com no mínimo 4Gb de RAM
- Os navegadores recomendados são Internet Explorer 11+, Edge 12+, Firefox 27+, Chrome 30+ ou Safari 7+
- Permissão de administrador da máquina
- Caso haja determinação legal para aplicação de avaliação presencial, ela será realizada em uma das unidades educacionais da FIA, em São Paulo/SP.

MATRIZ CURRICULAR

Os conceitos são apresentados por meio de exemplos e exercícios práticos monitorados pelo Professor.

CONTEÚDO PROGRAMÁTICO

> Introdução ao Spark e Python

- Visão geral do ecossistema Spark
- Componentes básicos do Spark
- Aplicações com Spark
- Estudo de caso
- Configuração do ambiente Spark

> Análise Exploratória de Dados

- Revisão dos principais conceitos do Python
- Introdução ao RDDs (Resilient Distributed Dataset)
- Transformações, Ações e DAG (Directed Acyclic Graph)
- API de programação RDD
- Realizar consultas interativas utilizando RDDs

> Spark SQL e DataFrame

- Conceitos de Spark SQL e DataFrame
- APIs DataFrame e SQL
- Otimização de consultas Catalyst
- ETL (Extrair, Transformar e Carregar)
- Realizando consultas com DataFrame e SQL
- Caching
- Visualização

> Spark Streaming

- Fontes de fluxos de dados em tempo real
- APIs disponíveis no Spark Streaming
- Confiabilidade e recuperação de falhas
- Otimização do desempenho
- Operações realizadas nos fluxos de dados

- Lendo dados de diferentes fontes (por exemplo: TCP, Kafka)
- Visualização contínua
- Visualizando os trabalhos (jobs) de fluxos na interface do Spark
- Extrair, transformar e carregar dados de múltiplas fontes de dados (JSON, base de dados relacionais, etc) com DataFrames
- Extrair dados estruturados de fontes de dados não estruturados através de transformações e ações
- Tratar valores faltando no conjunto de dados
- Aplicar as melhores práticas para análise de dados com Spark
- Realizar análise exploratória de dados utilizando DataFrames
- Visualizando dados através de bibliotecas populares de visualização do Python

> Funcionamento Interno do Spark

- Jobs, Stages e Tasks
- Desempenho do Job
- Uso de caching e melhores práticas
- Resolução de problemas utilizando a interface do Spark
- Visualizar como os jobs são divididos e executados dentro do Spark
- Análise dos logs dos executores
- Visualizar execução dos DAG (Directed Acyclic Graph)
- Visualizar consultas SQLs
- Observar a execução das tarefas (Tasks)
- Entendendo o desempenho
- Medindo a memória utilizada

> Introdução ao Aprendizado de Máquina

- Princípios básicos de Aprendizagem de Máquina
- Padrões de API de Aprendizagem de Máquina do Spark
- Utilizar algoritmos de aprendizagem de máquina através de pipelines e DataFrames
- Exemplos práticos para classificação e clusterização

COORDENAÇÃO



PROFA. DRA. ALESSANDRA DE ÁVILA MONTINI

Diretora do Labdata FIA, apaixonada por dados e pela arte de lecionar. Têm muito orgulho de ter criado na FIA cinco laboratórios para as aulas de Big Data e inteligência Artificial. Possui mais de 20 anos de trajetória nas áreas de Data Mining, Big Data, Inteligência Artificial e Analytics. Cientista de dados com carreira realizada na Universidade de São Paulo. Graduada e mestra em estatística aplicada pelo IME-USP e doutora pela FEA-USP. Com muita dedicação chegou ao cargo de professora e pesquisadora na FEA-USP, ganhou mais de 30 prêmios de excelência acadêmica pela FEA-USP e mais de 30 prêmios de excelência acadêmica como professora dos cursos de MBA da FIA. Orienta alunos de mestrado e de doutorado na FEA-USP. Membro do Conselho Curador da FIA, Coordenadora de Grupos de Pesquisa no CNPQ, Parecerista da FAPESP e Colunista de grandes Portais de Tecnologia.



[linkedin.com/in/alessandramontini](https://www.linkedin.com/in/alessandramontini)



Informações:

WhatsApp: (11) 95608-0719

labdata@fia.com.br / labdata.fia.com.br

UNIDADE PAULISTA - METRÔ BRIGADEIRO

Avenida Paulista, 302, 5º andar - CEP 01310-000 - Bela vista - São Paulo/SP

UNIDADE NAÇÕES UNIDAS

Avenida Doutora Ruth Cardoso, 7.221 - CEP 05425-070 - Pinheiros - São Paulo/SP