



# PÓS-GRADUAÇÃO

## ANALYTICS EM BIG DATA DATA ENGINEERING



360 HORAS



**lab.data**

Extensão, Pós-Graduação e MBA

# 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, Inteligência Artificial e Metaverso. É dirigido pela **Profª 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

+ 15 anos de atuação  
+ 15.000 alunos formados  
+130 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.



**Profª Drª 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.



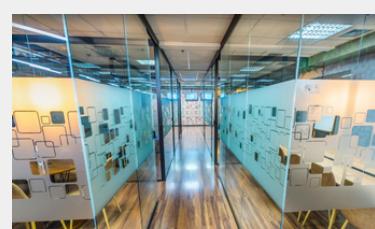
**Área para Networking**



**Espaços para Estudos**



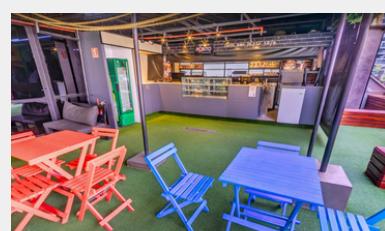
**Salas de Reunião**



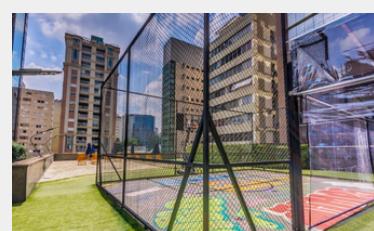
**Áreas de Descompressão**



**Espaço para Refeições**



**Quadra de Basquete e Praia**



# SOBRE O CURSO

## OBJETIVO

- Apresentar de forma clara, objetiva e com aplicações reais como as metodologias de Big Data e Inteligência Artificial podem ajudar as empresas a obterem vantagem competitiva;
- Apresentar como as tecnologias de Big Data e Inteligência Artificial podem ser utilizadas para a tomada de decisão e para a resolução de problemas complexos;
- Apresentar os principais componentes de Big Data utilizados em um projeto de engenharia de dados;
- Aplicar as tecnologias para tomada de decisão por meio de aulas teóricas, exercícios práticos, estudos de casos e projeto de Big Data.

## PERFIL DO ALUNO

Profissionais que manipulam e precisam tomar rápidas decisões por meio bases de dados e que tenham interesse em extrair informação por meio de modelos baseados em algoritmos de Inteligência Artificial e tecnologias de Big Data.

## 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.

# SOBRE O CURSO

## METODOLOGIA

- As aulas na modalidade presenciais serão realizadas em um ambiente propício para aprendizagem;
- As aulas na modalidade EaD ao vivo serão transmitidas por meio de uma plataforma digital;
- Todas as aulas terão a presença de um professor titular com apoio de pelo menos um professor assistente;
- Serão utilizados diversos recursos tecnológicos para o melhor entendimento dos conceitos;
- Todas as aulas serão gravadas e permanecerão disponíveis em nossa biblioteca digital até o final do curso.

## REQUISITOS PARA EAD AO VIVO

- 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 (13/i5/17 ou AMD equivalente) com no mínimo 16Gb 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.

# SOBRE O CURSO

## MATRIZ CURRICULAR

O curso será ministrado em dois módulos.

- No **primeiro módulo** serão apresentadas as tecnologias utilizadas por engenheiros e arquitetos de dados para a resolução de problemas complexos. Neste módulo será criada uma SOLUÇÃO ou PRODUTO baseado em dados que seja resiliente, escalável e confiável utilizando as principais tecnologias destacadas em aula;
- No **segundo módulo** serão apresentadas aplicações Analytics e Inteligência Artificial utilizadas para a tomada de decisão. Neste módulo serão implementadas algumas técnicas de Machine Learning e Deep Learning entregando valor ao projeto proposto;

**As informações estão sujeitas a alterações sem aviso prévio.**

# MÓDULOS

## 1º MÓDULO - TECNOLOGIA DE BIG DATA

Este módulo tem como objetivo apresentar, de forma prática, as principais tecnologias utilizadas nas organizações, por engenheiros e arquitetos de dados, para a implementação de projetos de Big Data e de Inteligência Artificial.

### > Introdução

- Conceito de Big Data, Inteligência Artificial, Machine Learning,
- Panorama para o surgimento do Big Data;
- Necessidades e vantagens da utilização das técnicas de Big Data e das metodologias de Inteligência Artificial;
- 7 Vs do Big Data: Variedade, Veracidade, Valor, Volume, Velocidade, Visualização e Vulnerabilidade;
- O profissional adequado para trabalhar com o Big Data: Data Scientist, Data Engineer e Data Architect;
- Cases nacionais e internacionais de Big Data.

### > Cloud Computing

- Fundamentos Cloud Computing;
- Cloud Adoption Framework;
- Conceitos e casos de uso de aplicações utilizando SaaS (Software-as-a-Service), Paas(Platform-as-a-Service) e IaaS (Infrastructure-as-a-Service);
- Finops;
- Ingestão de dados.

### > Programação

- Introdução a linguagem Python e suas principais bibliotecas;
- Análise de dados utilizando Python;
- Microservices em Python;
- Ingestão de dados em Python;
- Introdução a linguagem SQL;
- Conceitos de DML, DDL e DQL;
- Acesso a dados com linguagem SQL.
- Padrões de Microservices ;
- CI/CD;
- Observabilidade.

## > Plataformas de Dados

- Criação de plataforma de dados utilizando tecnologias de código aberto;
- Hadoop: Arquitetura e conceitos do framework pioneiro dos ecossistemas Big Data;
- Conceitos e utilização de conjuntos de práticas de automatização de processos DevOps;
- Introdução a Docker e containers e Kubernetes;
- Abordagem de arquitetura em micros serviços independentes;
- EDA - Conceitos da Arquitetura Orientada a Eventos e sua utilização com Streams de Eventos;
- MLOPS: Automatização de projetos de Machine Learning e suas práticas.

## > Armazenamento de Dados

- Utilização de dados estruturados, semiestruturados e não estruturados;
- Tipos de arquivos (CSV, JSON, Avro, Parquet, Delta e outros);
- Arquitetura e armazenamento de dados destruído com HDFS;
- LakeHouse: Transações ACID, Time Travel, Schema Evolution, Vacuum, Particionamento;
- Conceito, arquitetura e utilização de padrão S3 em cloud e on premise com Minio;
- Conceito, arquitetura, modelagem e utilização de banco de dados NoSQL das famílias Column Store (Cassandra), Document (MongoDB e Elastic), Key/Value (Redis), Neo4j (Grafo) e GEN IA Database (Vector);
- Arquitetura de dados relacional, internal engine e utilização de bancos de dados relacionais e Data Warehouse com MySQL e PostgreSQL.

## > Ingestão de Dados

- Fontes de dados e coleta de dados;
- Streaming de dados com Kafka Platform (Kafka, Kafka Connect, KSqlDb);
- Eventos em banco de dados utilizando CDC (Change Data Capture);
- Ingestão em tempo real e fluxo de dados com Kafka Connect, Spark, Nifi e AirByte;
- Ingestão batch, ETL e ELT utilizando Spark e AirFlow;
- Orquestração e deploy de pipeline de ingestão de dados.

## > Processamento e Análise de Dados

- Conceitos e Arquitetura Spark;
- Introdução ao processamento distribuído;
- Análise de dados com Python (PySpark) e SQL;
- Análise exploratória de dados com Spark;
- Processamento de dados em tempo real;
- Análise e cruzamento de dados em modo batch;
- Criação de aplicações distribuídas com Spark;
- Processamento ETL com DBT.

## > Acesso e Distribuição de Dados

- Conceito de acesso a arquivos com engine SQL, ClickHouse, Duckdb, Athena, Synapse, Trino, Presto e Dremio;
- Arquitetura e acesso a dados com Hive;
- Federação de dados e acesso In Memory com Trino/Presto;
- Entregando dados com API.

## > Governança

- Catálogo de metadados;
- Qualidade de dados;
- Linhagem;
- Gestão de acesso.

## > Transformação Digital

- Catálogo de metadados;
- Qualidade de dados;
- Linhagem;
- Gestão de acesso.

## > Projeto de Big Data

Será realizado um projeto de Big Data com as principais tecnologias utilizadas por engenheiros e arquitetos de dados.

O objetivo deste projeto é aplicar os conhecimentos adquiridos no curso para criar uma solução ou produto baseado em dados que seja resiliente, escalável e confiável utilizando as principais tecnologias de dados destacadas em aula. Para isso, é proposto desenhar e implementar uma arquitetura de dados com os seguintes pilares:

### **1. Coleta de Dados:**

- Implementar um sistema de ingestão de dados capaz de lidar com diferentes fontes, como bancos de dados, logs e streaming de dados em tempo real;
- Utilizar ferramentas como Plataforma Kafka, Spark Streaming, Airbyte, Nifi e Ariflow para garantir a escalabilidade e a confiabilidade na ingestão de dados.

### **2. Armazenamento:**

- Projetar e implementar um sistema de armazenamento utilizando tecnologias como HDFS e S3;
- Integrar bases de dados NoSQL, como Elastic ou MongoDB, para acesso de dados em baixa latência.

### **3. Processamento de Dados:**

- Aplicar técnicas de processamentos distribuídos utilizando Apache Spark e Python para realizar análises relevantes ao projeto;
- Implementar pipelines de processamento de dados eficientes para transformação e limpeza dos dados.

### **4. Análise e Visualização:**

- Disponibilizar acesso aos dados armazenados e processados com ferramentas de acesso como Presto, Hive, API, entre outros;
- Desenvolver dashboards interativos utilizando frameworks Metabase ou Apache Superset para fornecer insights açãoáveis aos usuários finais.

## **5. Segurança e Gerenciamento de Dados:**

- Implementar políticas de segurança para proteger dados sensíveis;
- Utilizar ferramentas de gerenciamento de metadados para rastrear a proveniência dos dados e garantir a conformidade com regulamentações.

## **6. Entregáveis:**

- Documentação técnica detalhada do sistema, abrangendo arquitetura, processos de ETL, segurança e manutenção;
- Código-fonte do sistema, hospedado em um repositório versionado (por exemplo, GitHub);
- Apresentação executiva destacando os principais resultados e benefícios obtidos com a implementação da solução.

## **2º MÓDULO - INTELIGÊNCIA ARTIFICIAL**

Este módulo tem como objetivo apresentar os fundamentos e a importância da inteligência artificial para a tomada de decisão.

### **> Inteligência Artificial**

- Introdução a inteligência artificial;
- Inteligência Artificial Aplicada aos Negócios;
- Importância da análise de imagem, som e texto.

### **> Análise Exploratória de Dados**

- Tipos de variáveis: dados qualitativos e quantitativos;
- Distribuição de frequências;
- Medidas de posição e dispersão;
- Gráficos: Box Plot e Histograma;
- Aplicações.

### **> Machine Learning**

- Aplicações com algoritmos de Machine Learning para tomada de decisão;
- Importância do aprendizado supervisionado e não supervisionado;
- Aprendizado Supervisionado - Problemas de Regressão - Modelagem e Pipeline de Dados
- Aprendizado Supervisionado - Problemas de Classificação - Modelagem e Pipeline de Dados
- Aprendizado Não Supervisionado - Problemas de Clusterização - Métricas para Avaliação de Modelos
  - Aplicações com os algoritmos:
    - Random Forests;
    - Gradiente boosting;
    - Bagging;
    - LGBM (LightGBM);
    - SVM (Support Vector Machine);
    - K-means.

### > Deep Learning

- Introdução a redes neurais;
- Importância da análise de imagem e texto;
- Aplicações com redes neurais densas, recorrentes, convolucionais;

Aplicações com Inteligência Artificial Generativa.

## PROJETO DE ANALYTICS

Neste projeto serão implementadas algumas técnicas de Machine Learning e Deep Learning entregando valor ao projeto de engenharia de dados proposto no módulo de tecnologias de big data.

## MÓDULO OPTATIVO - INTERNACIONAL

Explore novos horizontes!

Adquira habilidades em instituições internacionais por meio do nosso bloco optativo internacional.

Desperte para oportunidades inéditas e destaque-se como líder no mercado, se imergindo em experiências enriquecedoras em faculdades de renome internacional. Conheça nossas instituições conveniadas [aqui](#).

Amplie suas possibilidades e trilhe o caminho do sucesso global!

# DIREÇÃO



**PROF.ª DR.ª  
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://linkedin.com/in/alessandramontini)

# Informações

11 97132-3535

labdata@fia.com.br

labdata.fia.com.br

**UNIDADE PAULISTA - ESTAÇÃO BRIGADEIRO**

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

**UNIDADE NAÇÕES UNIDAS - ESTAÇÃO PINHEIROS**

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