



MBA em Engenharia de Dados

Estrutura Curricular – componente curricular/carga horária

Módulo 1 - Fundamentos de Dados e Analytics	
<i>Dados e Analytics nas Organizações</i>	32 horas
<i>Data Visualization</i>	32 horas
<i>Data Science Experience</i>	32 horas
Governança de Dados e Metadados	32 horas
<i>Data Lakes, Lakehouses e Data Meshes</i>	32 horas
Mentoring e Carreiras em Engenharia de Dados	32 horas
Hands-On Fundamentos de Dados e Analytics	24 Horas
Carga horária total do módulo	216 horas
Módulo 2 – Engenharia de Dados em Big Data	
Linguagens de Programação para Dados e Analytics (Python e R)	32 horas
<i>Data Collection & Storage</i>	32 horas
<i>Big Data Processing</i>	32 horas
<i>Data Prep & Transformation</i>	32 horas
DevOPS & DataOPS	32 horas
<i>Cloud Computing e SRE (Uma visão Prática)</i>	32 horas
Hands-On Engenharia de Dados em Big Data	24 Horas
Carga horária total do módulo	216 horas
Total da carga horária do curso	432 horas

IDENTIFICAÇÃO DA DISCIPLINA (01)

1. Nome do Componente Curricular: **Dados e Analytics nas Organizações**
2. Carga Horária: 32 h/aula
3. Ementa: Dados e *Analytics* como ativos estratégicos de valor nas organizações. Alinhamento estratégico das iniciativas de Dados e os objetivos organizacionais. Estruturas e organizações dos times de dados. Papéis e responsabilidades em Dados e *Analytics*.
4. Objetivo: Proporcionar aos alunos visões abrangentes de como Dados e *Analytics* podem contribuir para o sucesso das organizações, incluindo grandes transformações de negócios. Definição dos objetivos de desempenho e a criação



dos indicadores de desempenho (KPI's e OKRs). Explorar abordagens *Data Centric* e *Analytics by Design* onde dados permeiam toda a organização requerem estruturas distribuídas e federadas de dados que capacitam as organizações a desenvolverem diversas iniciativas de dados simultaneamente por toda a organização. Conhecer os diversos papéis desempenhados por profissionais de Ciência de Dados nas organizações. 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.

5. Conteúdo Programático:

- Dados e *Analytics* como elemento estratégico.
- Transformação de Negócios.
- Indicadores de desempenho (KPI's e OKRs).
- Abordagens *Data Centric*.
- *Analytics by Design*.
- Arquétipos de gestão Distribuída e Federada.
- Papéis e responsabilidades em Dados e *Analytics*.
- 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.



IDENTIFICAÇÃO DA DISCIPLINA (02)

1. Nome da Disciplina: **Data Visualization**
2. Carga Horária: 32 h/aula
3. Ementa: Exploração de diversas formas como os humanos percebem informação e podem gerar insights a partir de estímulos visuais. Como técnicas de visualização de dados podem apresentar indicadores de performance de negócios para tomada de decisão. Visualização de dados numéricos e não numéricos. O processo de criação e interpretação de visualização de dados. Exploração dos mais diversos tipos de representações gráficas de dados, como, tabelas, gráficos, mapas, etc. *Storytelling* com Dados
4. Objetivo: Apresentar diferentes e efetivas formas de representar graficamente dados numéricos e não numéricos para geração de insights de negócios que suportam a tomada de decisão. Dar ao aluno condições de escolher a forma mais apropriada para apresentar informações visualmente, da forma mais adequada ao público e a mensagem que se deseja transmitir. Apresentar ao aluno ferramentas que facilitem o processo de visualização de dados.
5. Conteúdo Programático:
 - Representação visual em Tabelas, gráficos e mapas
 - O ser humano: memória, retina e raciocínio.
 - Dados, mapeamento e cartas.
 - Coordenadas paralelas e empilhamento de gráficos.
 - Usando cores e as regras de Tufte.
 - Sistemas geolocalizados.
 - Árvores de mapas e escalas multidimensionais.
 - Sistemas de visualização.
 - Visualização de informação e de bases de dados.
 - Projeto de sistemas de visualização.



IDENTIFICAÇÃO DA DISCIPLINA (03)

1. Nome da Disciplina: ***Data Science Experience***
2. Carga Horária: 32 h/aula.
3. Ementa: Explorar os fundamentos de Ciência de Dados com base em Análise Estatística e Aprendizagem de Máquina na prática.
4. Objetivo: Estudar os conceitos e práticas da análise estatística de dados, apresentando ferramentas e técnicas que possibilitem a identificação ou predição de fenômenos quantificáveis. Entender os conceitos de Aprendizagem de Máquina. Entender os principais tipos de aprendizagem. Conhecer as principais tarefas que podem ser realizadas com Aprendizagem de Máquina. Explorar Técnicas e ferramentas para geração de modelos de Data Science automáticos (Auto-ML),
5. Conteúdo Programático:
 - Preparação de Dados (escala de medidas, variáveis).
 - Análise Exploratória de Dados (tipos de tabulação e representação gráfica).
 - Distribuições de Probabilidade.
 - Teste de Hipóteses.
 - Análise de Regressão.
 - Análise Preditiva: análise de categorias (saída discreta/categórica) e previsão de séries temporais (saída contínua).
 - O paradigma de Aprendizagem de Máquina.
 - Aprendizagem não supervisionada, supervisionada e por reforço.
 - 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.
 - Ferramentas para Auto-ML



IDENTIFICAÇÃO DA DISCIPLINA (04)

1. Nome da Disciplina: **Governança de Dados e Metadados**
2. Carga Horária: 32 h/aula.
3. Ementa: Governança de Dados como acelerador de iniciativas de transformação de negócios. Qualidade de Dados. Risco, Segurança, Privacidade e Compliance de dados. Pessoas e processos de Governança de Dados. Gestão de Metadados. Catálogo de Dados como fonte de busca de informação.
4. Objetivo: Explorar os conceitos fundamentais da governança de dados e como sua boa estruturação pode contribuir com as principais iniciativas de transformação de uma organização. Entender os principais processos que podem contribuir com a governança de dados. Capacitar os participantes na criação de políticas que garantam a segurança e privacidade dos dados. Estudar os princípios da gestão de metadados e o papel do Catálogo de Dados em uma organização.
5. Conteúdo Programático:
 - Fundamentos de Governança de Dados;
 - Frameworks e Melhores Práticas para Governança de Dados;
 - Governança de Dados como um ativo organizacional;
 - Governança de Dados apoiando ações de transformação;
 - Frameworks e Processos de Governança de Dados;
 - Qualidade de Dados;
 - Risco, Segurança, Privacidade e Compliance de dados;
 - Lei Geral de Proteção de Dados (LGPD);
 - Gestão de Metadados;
 - Catálogo de Dados.



IDENTIFICAÇÃO DA DISCIPLINA (05)

1. Nome da Disciplina: ***Data Lakes, Lakehouses e Data Meshes***
2. Carga Horária: 32 h/aula
3. Ementa: Arquiteturas para ambientes de *Big Data*, incluindo *Data Lakes, Lakehouses e Data Meshes*. Visão arquitetural de Coleta, Armazenamento, Processamento e Serviço/Distribuição de Dados em múltiplas latências e tempestividades. Componentes de Autenticação e Autorização em ambientes de *Big Data* e Orquestração/Agendamento de rotinas. Evolução de *Dawarehouses, Data Marts* para arquiteturas de *Big Data*
4. Objetivo: Apresentar os principais componentes da Arquitetura de *Big Data*, as soluções de *Data Lakes, Lakehouses e Data Meshes* e como os dados são organizados com uma visão de evolução arquitetural de dados.
5. Conteúdo Programático:
 - Desenhos Arquiteturais Conceituais, Lógicos e Físicos.
 - Arquiteturas de *Data Lakes, Lakehouses e Data Meshes*.
 - Ferramentas de *Big Data*.
 - Visão arquitetural de Coleta, Armazenamento, Processamento e Serviço/Distribuição de Dados.
 - Visão de organização dos dados com abordagem de evolução arquitetural (Dados Brutos, Dados Refinados e Dados Especializados).
 - Autenticação e Autorização.
 - Orquestração e Agendamento de rotinas.



DENTIFICAÇÃO DA DISCIPLINA (06)

1. Nome da Disciplina: **Mentoring e Carreiras em Engenharia de Dados**
2. Carga Horária: 32 h/aula
3. Ementa: oportunizar aos estudantes conhecer práticas e desafios voltados ao desenvolvimento da carreira atual, assim como desenvolver uma visão crítica sobre as características e os problemas específicos contemporâneos, além de aprofundar o conhecimento sobre carreira e carreira empreendedora como alternativa de mercado de trabalho. O programa foi criado a partir da observação e da identificação de algumas lacunas que o currículo acadêmico se propõe contribuir.
4. Objetivo: Apoiar o educando para planejar sua carreira através do autoconhecimento, clareza e segurança ao fazer escolhas e tomar decisões cruciais relativas à sua vida profissional. Desenvolver novas competências socioemocionais que promovam desenvolvimento da carreira do aluno. Ajudar o aluno a superar bloqueios e barreiras que impeçam ou dificultem seu crescimento profissional.
5. Conteúdo Programático:
 - Desafios da carreira, linha do tempo da carreira e suas necessidades
 - Estágios da carreira
 - Tipos de carreira
 - 7 estágios da carreira
 - Carreira e sucessão
 - Mapeamento Pessoal de motivadores, interesses e talentos
 - Perfil do empreendedor e o mercado da tecnologia
 - Princípios do Empreendedorismo de carreira
 - Âncoras de carreira
 - Fases da construção da carreira
 - Inventário de atitudes e carreira
 - Perfil do empreendedor de carreira
 - Comportamento do empreendedor na carreira
 - Redes sociais na empregabilidade



- Construção do LinkedIn e CV
- Inventário e simulado de entrevistas
- Drivers do século XXI
- Mundo Vuca e Bani no empreendedorismo de carreira

IDENTIFICAÇÃO DA DISCIPLINA (07)

1. Nome da Disciplina: **Hands-On Fundamento de Dados e Analytics**
2. Carga Horária: 24 h/aula
3. Ementa: Aspectos práticos de Governança e Visualização de dados, análises preditivas e prescritivas e implementações de experimentos práticos baseados em problemas análogos aos encontrados nas organizações.
4. Objetivo: Capacitar o aluno nas principais ferramentas, práticas e técnicas do mercado de Dados & Analytics.
5. Conteúdo Programático:
 - Ferramentas de visualização de Dados;
 - Ferramentas de MLOps;
 - Ferramentas de AutoML;
 - Ferramentas para Governança de Dados.

IDENTIFICAÇÃO DA DISCIPLINA (08)

1. Nome da Disciplina: **Data Collection & Storage**
2. Carga Horária: 32 h/aula
3. Ementa: Processo de aquisição de dados. Modalidades de coleta: manual, automatizada, em lote, em fluxo contínuo, *Change Data Capture (CDC)*, etc.. Coleta por meio de sensores, Coleta por meio de conectores. Pré-processamento de dados. Persistência em *Clusters* e *Object Storage*, formatos de Armazenamento e



compactação de dados. Ciclo de vida e arquivamento de dados. Paradigmas de persistência: Relacional, Chave-Valor, Orientado a Documentos, Família de Colunas e Grafos.

4. Objetivo: Capacitar o aluno para a prática de coleta de dados em lote ou em fluxo contínuo, 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 *object storage* e os paradigmas de persistência que apoiam aplicações de *Big Data*, em termos de volume, velocidade e variedade. Explorar melhores estratégias de formato de armazenamento, serialização, compactação, etc. Estudar estratégias de ciclo de vida dos dados e arquivamento.

5. Conteúdo Programático:

- Processo de coleta e aquisição de dados;
- Modalidades e mecanismos de coleta de coleta;
- Pré-processamento;
- Armazenamento em *Cluster* e *Object Storage*;
- Estratégias de ciclo de vida dos dados e arquivamento;
- Formatos de arquivos, SerDe, Compactação;
- Leitura de arquivos compactados;
- Paradigmas de persistência;
- Formatos Open Data (Delta Lake, Apache Iceberg e Apache Hudi)
- Tendências em coleta e atuação. Cases e exemplos.



IDENTIFICAÇÃO DA DISCIPLINA (09)

1. Nome da Disciplina: ***Big Data Processing***
2. Carga Horária: 32 h/aula.
3. Ementa: Processamento em Big Data. Ambientes distribuídos e centralizados. Arquiteturas de processamento. Processamento distribuído. Processamento de dados heterogêneos. HPC Cluster Computing, incluindo Processamento Massivamente Paralelo (MPP) e Multiprocesadores Simétricos (SMP). Características de processamento MapReduce e Spark. Processamentos de Arquivos e Dataframes em Big Data.
4. Objetivo: Proporcionar aos participantes detalhamento do processamento de dados em ambientes *Big Data*, abordando suas arquiteturas e características. Capacitar os envolvidos a atribuírem arquiteturas de processamento para *Big Data*. Explorar melhores práticas e conceitos de dimensionamento de capacidade computacional para processamento para Big Data.
5. Conteúdo Programático:
 - Computação Distribuída;
 - Computação Paralela;
 - Processamento síncrono e assíncrono;
 - Escalabilidade Vertical e Horizontal;
 - Compartilhamento de memória;
 - Projetando capacidade computacional de processamento;
 - Processamento MapReduce e Spark;
 - Processamentos de Arquivos e Dataframes;
 - Processamento de dados não estruturados.

IDENTIFICAÇÃO DA DISCIPLINA (10)

1. Nome da Disciplina: ***Data Prep & Transformation***
2. Carga Horária: 32 h/aula.
3. Ementa: Transformação de Dados Transacionais Brutos em Dados Especializados para *Analytics*. Converter dados de diversos formatos e origens (estruturadas e não estruturadas) em estruturas prontas para análise. Abordagem de Modelagem de Dados e evolução arquitetural dos dados durante o fluxo analítico.



4. Objetivo: Capacitar os participantes na preparação de dados para analytics, entendendo as necessidades de dados, elencar as fontes de dados necessárias e realizar os processos de transformação de dados necessárias para consumo.

5. Conteúdo Programático:

- Modelagem de Dados Transacional Relacional Normalizada e NoSQL.
- Modelagem de Dados para Analytics Star Schema, Snow Flake, Data Vault, Flat Table, Wide Table, NoSQL, etc.
- Processos de Transformação de Dados Brutos em Dados Especializados para análise.
- Criação de Rotinas de Transformação por código, SQL, Low-Code/No-Code.
- Tempestividades de transformação Batch, Micro-Batch e Fluxo Contínuo.
- Limpeza e enriquecimento de dados. Manipulação, tratamento e transformação de dados não estruturados e semiestruturados.

IDENTIFICAÇÃO DA DISCIPLINA (11)

1. Nome da Disciplina: **DevOPS & DataOPS**

2. Carga Horária: 32 h/aula.

3. Ementa: Conceitos de Engenharia de Software. Práticas e Técnicas de *Continuous Integration and Continuous Delivery*. Observability e Monitoramento de Aplicações. Como DataOps acelera o fluxo de desenvolvimento de aplicações para manipulação de dados. Cultura DevOps e Framework SRE

4. Objetivo: Conhecer os principais conceitos de Engenharia de Software. Capacitar os participantes na criação de esteiras de código para manipulação de dados. Explorar conceitos de monitoramento de aplicações. Apresentar as características da Cultura DevOps.

5. Conteúdo Programático:

- CALMS: Pilares do DevOps;
- Ciclo de Vida de Aplicações;
- Shift Left;
- Processos de Desenvolvimento de Software;
- Métricas de produtividade. *Metricas Accelerate (Deployment Frequency (DF), Lead Time to Changes (LTTC), Mean Time To Recovery (MTTR) and Change Failure Rate (CFR))*;



- Padrões de Código e Style Guides;
- Processos de Revisão de Código;
- Processos de Documentação;
- Práticas de Teste de Software;
- Práticas e processos de versionamento de código;
- Repositórios de Código;
- Ferramentas e *Pipelines* para *Continuous Integration*, *Continuous Deployment* e *Continuous Delivery*;
- Observability e Monitoramento de Aplicações;
- DataOPS e manipulação de dados.

IDENTIFICAÇÃO DA DISCIPLINA (12)

1. Nome da Disciplina: **Cloud Computing e SRE (Uma visão Prática)**
2. Carga Horária: 32 h/aula.
3. Ementa: Nessa disciplina serão debatidos os fundamentos da computação em nuvem, detalhados os modelos de serviços em nuvem, incluindo infraestrutura como serviço, software como serviço e plataforma como serviço em uma visão prática focada em Big Data & Analytics, com foco em manter ambientes plataformas de dados confiáveis e resilientes em ambientes de nuvem.
4. Objetivo: Capacitar no desenho para implementação de ambientes e plataformas de dados em nuvem
5. Conteúdo Programático:
 - Introdução ao Cloud Computing
 - Conceitos básicos e terminologias
 - A importância do Cloud no cenário atual
 - Modelos de Negócios na Nuvem
 - Conceitos e Definição de IaaS, Paas e SaaS
 - Principais Provedores
 - Infraestrutura como código (Terraform)
 - Ferramentas de Big Data na Nuvem
 - SRE: Engenharia de Confiabilidade do site
 - Métricas SRE (SLI's, SLA's e SLO's e Error Budgets)



IDENTIFICAÇÃO DA DISCIPLINA (13)

1. Nome da Disciplina: **Linguagens de Programação para Dados e Analytics (Python e R)**
2. Carga Horária: 32 h/aula
3. Ementa: Aspectos básicos de programação das linguagens Python e R para Ciência de Dados. Programação de aplicações, apresentação e desenvolvimento de casos.
4. Objetivo: Capacitar o aluno a ler, compreender e adaptar 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.
5. 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. Tidy 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.

IDENTIFICAÇÃO DA DISCIPLINA (14)

1. Nome da Disciplina: **Hands-On Engenharia de Dados em Big Data**
2. Carga Horária: 24 h/aula
3. Ementa: Aspectos práticos de Engenharia de dados, incluindo transformação de dados, implementações de experimentos práticos baseados em problemas análogos aos encontrados nas organizações.
4. Objetivo: Capacitar o aluno nas principais ferramentas, práticas e técnicas do mercado de Dados & Analytics, incluindo abordagens em cloud computing de Data Lakes.
5. Conteúdo Programático:
 - Ambientes de Cloud Computing
 - Ferramentas de Ingestão de Dados
 - Ferramentas de Armazenamento de Dados
 - Ferramentas de Transformação de Dados
 - Ferramentas de Serviço e Distribuição de Dados