

# Nome do Curso: Engenharia de Dados Online

- Justificativa: O mundo dos negócios tem-se transformado com soluções inovadoras baseadas em dados. As possibilidades de uso dos dados, a sua gestão e as tecnologias de apoio ao tratamento de dados já têm um olhar estratégico por parte das organizações. Percebe-se que o dia-a-dia de uma economia altamente competitiva e globalizada tem posto como fator crítico de sucesso e gerador de diferenciais competitivos para o negócio a capacidade das organizações em ter acesso e poder analisar, no tempo certo e de forma segura, seus dados, informações e conhecimentos necessários para a sua gestão. A evolução da tecnologia de gerenciamento de dados vem sendo acompanhada de avanços teóricos e conceituais importantes, que permitiram desenvolver uma ampla variedade de novas aplicações, métodos e técnicas de trabalho. O gerenciamento de dados é hoje utilizado não apenas em sistemas de informação convencionais, mas também em ambientes distribuídos em rede, na descoberta de conhecimento para tomada de decisão, no tratamento de dados geoespaciais e em diversas outras áreas. Essas novas possibilidades influenciam de forma decisiva a maneira segundo a qual acumulamos dados e os transformamos em informação. Com o enorme volume e variedade de dados disponíveis atualmente e com uma ampla oferta de tecnologias, as organizações buscam cada vez mais profissionais capazes de gerenciar esta quantidade massiva de dados, utilizando-se de novas ferramentas e técnicas para administrar essa complexidade. Esses profissionais devem ter formação em Engenharia de Dados, que é uma profissão em ascensão preocupada com a integração de tecnologias emergentes e com abordagens de big data para aquisição, gerenciamento e análise de dados. O profissional engenheiro de dados é responsável pela criação de soluções para processar e tratar dados para aplicações que demandam dados, especialmente para o grande volume e variedade de dados disponíveis atualmente (Big Data). Esse profissional utiliza conhecimentos em ciência da computação para criar sistemas e resolver problemas de processamento de dados em tempo real e manipular quantidade massivas de dados. Ele deve ter experiência com banco de dados relacionais e não relacionais, e saber como projetar e implementar arquiteturas escaláveis, confiáveis e eficientes para fornecer, monitorar e gerenciar todo o ciclo de vida dos dados. Nesse cenário é importante perceber que o conhecimento em poder dos profissionais da área torna-se obsoleto com muita velocidade. Esses profissionais precisam manter-se atualizados o tempo todo, uma vez que o mercado de trabalho vai sendo alterado em função da própria evolução tecnológica, relegando à obsolescência aqueles que não puderem (ou não conseguirem) acompanhar os processos de mudança. Observa-se também que engenharia de dados é uma das ocupações de tecnologia que mais cresce, onde a demanda por engenheiros de dados qualificados supera em muito a oferta. Apesar da alta demanda no mercado, mídias especializadas em gestão de carreiras reforçam que existe uma notória carência de profissionais experientes com esse perfil. De acordo com o “Relatório Anual de Posições Emergentes” do LinkedIn, a Engenharia de Dados é uma das dez profissões com maior potencial de crescimento nos próximos anos. Nesse contexto, o curso de Especialização em Engenharia de Dados pretende oferecer, além de uma visão abrangente das tecnologias, dos produtos e das áreas de aplicação de Engenharia de Dados, uma sólida base conceitual que permita não apenas a aquisição contínua de conhecimentos atualizados, mas também a absorção,

com maior facilidade, de futuras evoluções tecnológicas. Além disso, promover o entendimento sobre os variados tipos de tecnologia aplicada para um mercado com alta demanda de pessoas com capacidade de gerar valor para as empresas. Ou seja, formar engenheiros de dados capazes de fornecer uma abordagem ampla no projeto, implementação e governança de uma arquitetura de dados plenamente aderente ao negócio. Matriz curricular atualizada para atender as necessidades do mercado, oferecendo uma formação focada na definição de arquiteturas de dados em tecnologias atuais; Conteúdos apresentados por meio de casos reais, que colocam o aluno próximo de situações comuns no dia a dia de um usuário de dispositivo móvel; Tradição de ensino PUC Minas; Professores com muita experiência de mercado e com uma sólida formação acadêmica; Mentores experientes focados em orientar e motivar para otimizar o aprendizado; Abordagens inovadoras de ensino-aprendizagem em que as aulas e atividades pedagógicas são centradas nas necessidades dos alunos. Elas seguem dinâmicas orientadas por princípios de metodologias ativas; Experiência de aprendizado é suportada por ferramentas interativas - acessível via Web ou dispositivos móveis - incluindo salas virtuais, bate-papos e fóruns de discussão para estimular o aluno a um maior engajamento com o seu curso; Aprendizagem flexível em que o aluno planeja o próprio ritmo para alcançar seus objetivos pessoais;

- Objetivos: Definir conceitos relacionados a banco de dados e aos processos e tecnologias que orientam a definição da arquitetura de dados e verificar seus impactos em novos projetos; Propor, planejar e desenvolver projetos de sistemas de bancos de dados; Planejar e definir recursos que envolvam arquiteturas de dados altamente escaláveis; Criar e usar soluções com Python e SQL; Construir soluções que suportem a aquisição, transformação e o consumo dos dados; Construir e otimizar pipelines de dados, enfileiramento de mensagens e processamento de streaming de dados; Identificar alternativas para garantir e melhorar a confiabilidade, a qualidade e a segurança dos dados; Definir o processo de entrega e integração contínua e utilizar ferramentas para a gestão e controle dos ativos de uma arquitetura de dados; Prospectar novas tecnologias para gestão de dados e auxiliar na sua incorporação às estratégias, planejamentos e práticas da organização; Atuar de forma empreendedora na criação de novos negócios, utilizando os conceitos e as técnicas aprendidos no curso. O especialista em Engenharia de Dados poderá atuar como Engenheiro de Dados nas mais diversas áreas que desenvolvem soluções baseadas em dados ou em iniciativas para dar suporte aos mais variados produtos digitais e em diversos papéis como: Arquiteto de Banco de Dados, DBA, Analista e Cientista de Dados, Gerente de Projetos de Software, entre outros. Além disso, ele poderá atuar nos variados tipos de projetos inovadores em TI.
- Público Alvo: Profissionais com formação superior: Em Engenharia de Software, Sistemas de Informação, Ciência da Computação, Engenharia de Computação e tecnólogos da área de Tecnologia da Informação e outros cursos correlatos; Nas mais diversas áreas e que tenham alguma experiência em desenvolvimento de software; Em busca de novas habilidades, soft skills e networking em Engenharia de Dados.

## Disciplinas:

Disciplina 1: PROCESSAMENTO DE FLUXOS DISCRETOS E

## **CONTÍNUOS DE DADOS**

Ementa: Processamento em lotes. Processamento de dados gerados por eventos. Spark, spark-streaming, Kafka, serviços gerenciados em cloud, soluções serverless e orientadas a eventos. Catalogação de streaming para controle de jobs. Configuração do Spark.

## **Disciplina 2: ARQUITETURAS E SERVIÇOS DE DATA LAKES E DATA WAREHOUSING**

Ementa: Conceitos de Data Lake e Data Warehouse. Dados Estruturados x Dados Não Estruturados. Logical Data Lake. Data ponds. Blocos funcionais em uma arquitetura de dados. Schema on-write x Schema on-read. Sincronização de fluxos independentes e dependentes. Arquitetura Corporativa - Enterprise Data Hub. Construção de arquiteturas com alta disponibilidade (HA). Construção de dimensões em DW. Engines de Data Warehousing. Orquestração de transformações em DW. Organização de camadas em um Data Lake. Data Lake On-premise x Data Lake na Nuvem.

## **Disciplina 3: PYTHON PARA ENGENHARIA DE DADOS**

Ementa: Fundamentos da Linguagem Python. Bibliotecas para manipulação de dados. Bibliotecas para visualização de dados. Conexões com Bancos de Dados e arquivos. Construção de módulos de migração. Operações de processamento e extração. Manipulação de dados com APIs. Implementação de módulos serverless. Processamento multi-threading e distribuídos.

## **Disciplina 4: INGESTÃO E CATALOGAÇÃO DE DADOS**

Ementa: Conceitos e técnicas de ingestão de dados. Ingestão Change Data Capture (CDC). Slowly Change Dimension. Técnicas de ingestão baseadas em diferenças temporais. Conectores de migração em tempo-real. Ingestão por ETL / ELT clássico. Fluxo de dados real Time e CEP (Complex Event Processing) Tratamento de eventos. Mecanismos de inferência de tipagem. Técnicas de consistência e harmonização de tipagem. Schema assert. Schema-registry. Monitoramento de camadas de ingestão de dados. Organização de áreas de dados RAW. Ferramentas e tecnologias para ingestão de dados. Setup de infraestrutura de rede.

## **Disciplina 5: OTIMIZAÇÃO, MONITORAMENTO E OPERAÇÃO EM BANCOS DE**

Ementa: Indexação. Análise de desempenho. Otimização e tuning (relacional e NoSQL). Benchmarks para SGBD. Construção de soluções de monitoramento de arquiteturas de dados. Indicadores, alarmes e métricas de acompanhamento. Serviços de alertas baseados em cloud. Recuperação de falhas. Sistemas de backup. Bancos de dados ativos (triggers). Ferramentas de visualização para monitoramento.

## **Disciplina 6: SEGURANÇA EM ARMAZENAMENTO, TRATAMENTO E CONSUMO DE DADOS**

Ementa: Criptografia e gestão de chaves criptográficas. Serviços e técnicas de autenticação. Tratamento de dados sensíveis e mascaramento. Planejamento e regras de rede. Gestão de acessos de usuários e serviços. Implementação de padrões e processos compatíveis com LGPD / GDPR. Análise de riscos. A necessidade da adoção dos controles internos. Conceitos de auditoria.

## **Disciplina 7: BANCOS DE DADOS NOSQL**

Ementa: Motivações para bancos NoSQL. Conceitos e características dos bancos de

dados NoSQL Principais benefícios de um banco de dados NoSQL. Principais tipos de Bancos NoSQL. Key-value databases. Document databases. Column Family Databases. Graph Databases. Principais diferenças arquiteturais entre os tipos. Principais casos de uso de cada tipo.

## **Disciplina 8: ARMAZENAMENTO E PROCESSAMENTO MASSIVO E DISTRIBUÍDO DE DADOS**

Ementa: Introdução aos sistemas distribuídos. Princípios de processamento e de volumes de dados massivos. Sistemas de arquivos distribuídos. Modelo MapReduce. Balanceamento de carga. Replicação e redundância de dados e metadados. Gestão de metadados. Sincronização e disponibilidade. Operações de leitura, escrita e listagem de arquivos distribuídos. Soluções em Processamento Distribuído. Técnicas de compressão, otimização e particionamento de dados.

## **Disciplina 9: PREPARAÇÃO, ORQUESTRAÇÃO E FLUXOS DE DADOS**

Ementa: Montagem do conjunto de dados. Tipos de dados. Melhoramento e enriquecimento. Preparação: eliminação de dados irrelevantes, granulação e agregação, consistência, poluição, análise de domínios de atributos, integridade, concordância, duplicação e redundância. Feature Engineering. Combinando dados de múltiplas fontes. ELT x ETL. Transformação e transferência de dados. Ferramentas para preparação de dados. Ferramentas de orquestração. Escalonamento de jobs baseados em eventos. Reprocessamento em serviços de migração. Resiliência em migração. Orquestradores open-source. Orquestradores baseados em cloud. Implementação de fluxos de correção de dados em expurgo.

## **Disciplina 10: CAMADAS E SERVIÇOS DE CONSUMO DE DADOS**

Ementa: Produtores e consumidores de dados. Dados de domínio como produto. Camadas de dados. Critérios para determinar a camada de dados. Barramentos de mensageria de dados. Particionamento e técnicas de cache para acessos de alta frequência. Planos de execução. Consumo de alta volumetria de dados. Camadas de consumo por serviços. Autenticação e controle de acessos. Design convergente para dados e plataforma self-service.

## **Disciplina 11: GOVERNANÇA DE DADOS**

Ementa: Contexto organizacional de dados. Conceitos de Governança de Dados (GD). Framework DMBOK. Políticas, padrões e procedimentos aplicados aos dados: Data Stewardship, Data Owners, Dados Mestres, Dados Referência, Metadados, Data Catalog. Processo de implantação de GD. Modelos de maturidade de dados. GD aplicada em leis de Proteção (LGPD-GDPR). Compliance e Risk Assessment. GD 2.0: Ética nos dados, Agilidade em GD, Gerência de Mudanças.

## **Disciplina 12: HUMANIDADES**

Ementa: O ser humano, o processo de humanização e o conceito de pessoa. Desafios contemporâneos e o lugar da religião e da espiritualidade. Autonomia e heteronomia na sociedade atual. Princípios éticos e ética profissional.

## **Disciplina 13: COMPUTAÇÃO EM NUVEM**

Ementa: Aspectos da Computação em Nuvem: conceitos, tipos, utilização, fornecedores, utilização corporativa. Soluções corporativas de SaaS, PaaS e IaaS. Aspectos de segurança da informação. Principais plataformas de Colaboração e Comunicação. Arquiteturas de Cloud: MultiCloud e Cloud híbridas.

#### **Disciplina 14: BANCOS DE DADOS RELACIONAIS E LINGUAGEM SQL**

Ementa: Fundamentos de bancos de dados. Modelagem de bancos de dados relacionais (Diagrama ER). Linguagem SQL: Data Manipulation Language (DML) e Data Definition Language (DDL).

#### **Disciplina 15: CULTURA E PRÁTICAS DEVOPS**

Ementa: A cultura DevOps. Integração contínua e entrega contínua. Estratégias de deploy. Projeto de pipeline para build e deployment. Automação de testes. Infrastructure as Code (IaC). Ferramentas e infraestrutura do ambiente integrado DevOps: Containers, Docker, Kubernetes e OpenShift.

#### **Disciplina 16: CULTURA E PRÁTICAS DATAOPS E MLOPS**

Ementa: A cultura DevOps. Integração contínua e entrega contínua. Estratégias de deploy. Projeto de pipeline para build e deployment. Automação de testes. Infrastructure as Code (IaC). Ferramentas e infraestrutura do ambiente integrado DevOps: Containers, Docker, Kubernetes e OpenShift.