

# Nome do Curso: Arquitetura de Dados

- 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 e em diversas outras áreas. Essas novas possibilidades influenciam de forma decisiva a maneira segunda a qual acumulamos dados e os transformamos em informação. O conceito de gestão de dados e as novas oportunidades profissionais a ela vinculadas surgiram como consequência da proliferação das tecnologias da informação e da geração massiva de dados. O desafio que as empresas têm hoje é ter uma política de governança de dados que facilite a gestão e análise desses dados para a tomada de decisões estratégicas na condução dos mais variados projetos. Nesse sentido, é importante considerar o papel do profissional Arquiteto de Dados. Esses profissionais são responsáveis por manter as funcionalidades dos bancos de dados da organização. Devem ter formação em Arquitetura de Dados, que é uma profissão em ascensão, preocupada, segundo Giovani Stefani (LinkedIn), com um conjunto de regras, políticas, padrões, modelos que governam e definem o tipo de dados coletados e como eles são usados, armazenados e integrados em uma organização e em seu sistema de banco de dados. Os Arquitetos de Dados traduzem requisitos de negócios para bancos de dados em requisitos técnicos, definindo princípios e padrões de dados. Eles devem gerenciar, explorar, analisar e tomar decisões apropriadas de gerenciamento de dados com base na sua geração, captura e processamento. Devem também saber projetar e implementar arquiteturas escaláveis, confiáveis e eficientes para fornecer, monitorar e gerenciar todo o ciclo de vida dos dados. 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. Observa-se também que Arquitetura de Dados é uma das ocupações de tecnologia que mais cresce e que a demanda por profissionais 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. Ofertas de emprego disponíveis, por exemplo no LinkedIn e Glassdoor, tem aumentado consideravelmente e de forma sustentável. Segundo dados do CAGED, nos últimos anos houve um aumento de 11,29% nas contratações formais com carteira assinada em regime integral de trabalho desse profissional. Nesse contexto, o curso de Especialização em Arquitetura de Dados pretende oferecer, além de uma visão abrangente das tecnologias, dos produtos e das áreas de aplicação de Arquitetura 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 Arquitetos 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 software 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; Professores 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: O curso de Arquitetura de Dados tem como objetivo principal capacitar profissionais com o perfil de Arquiteto de Dados com competências adequadas para promover a cultura de dados e para desenvolver projetos para armazenamento, tratamento e disponibilização de dados necessários a organização. Ao final do curso, o aluno será capaz de: Entender a importância da cultura de dados e aplicar metodologias para desenvolver estratégias de dados alinhadas à visão da organização e para criar valor para as soluções digitais; Conhecer sobre os processos, frameworks e tecnologias que orientam a definição da arquitetura de infraestrutura escalável para gerenciamento de dados; Projetar, desenvolver e gerenciar projetos que demandam técnicas e tecnologias atuais em arquitetura de dados e engenharia de dados; Conhecer técnicas, ferramentas e tecnologias para promover a gestão do ciclo de vida do dado, observando os processos de sua criação, aquisição, integração, gestão e operacionalização; Descrever os requisitos para garantir a confidencialidade, integridade e disponibilidade das informações bem como desenvolver procedimentos para gerenciamento de incidentes de dados; prospectar tendências na área de arquitetura e engenharia de dados.
- Público Alvo: Profissionais com formação superior: Engenharia de Software, Sistemas de Informação, Ciência da Computação, Engenharia de Computação e tecnólogos da área de Tecnologia da Informação e outros cursos correlatos; Em busca de novas habilidades, soft skills e networking em Arquitetura e Engenharia de Dados. Que já atuam em projetos ou atividades relacionados à análise e gestão de dados e que queiram complementar e aperfeiçoar seus conhecimentos em arquitetura de dados; Nas mais diversas áreas e que tenham alguma experiência em desenvolvimento de software; Que atuam ou pretendem atuar no desenvolvimento de soluções inovadoras e inteligentes para favorecer a transformação digital.

**Disciplinas:**

## **Disciplina 1: 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 2: BANCOS DE DADOS RELACIONAIS E NÃO RELACIONAIS**

Ementa: Modelo Relacional. SQL. Bancos de Dados NoSQL: definição; motivação; modelo de Transações. Modelos NoSQL. Propriedades Modelo Relacional x Propriedades Modelos NoSQL. Principais SGBD's.

## **Disciplina 3: ESTRATÉGIA DE DADOS**

Ementa: Fundamentos. Necessidades estratégicas de dados. Empresas Orientadas a Dados. Cultura Data Driven. Data literacy. Produtos de dados. Topologias produtos de dados. Principais componentes de uma estratégia de dados. Métodos para criação de estratégia de dados. Definição e execução da estratégia de dados. Estratégia de dados para tomada de decisão. Design de produtos de dados. Avaliação do ROI da estratégia de dados. Estudo de Casos.

## **Disciplina 4: MONITORAMENTO E OBSERVABILIDADE**

Ementa: Processo de tomada de decisão. Monitoramento x Observabilidade. Elementos, pilares e benefícios da observabilidade. Estratégias para medições e monitoramento contínuo. Conexão do monitoramento e observabilidade com as estratégias de SLO e Error Budgeting. Principais ferramentas de monitoramento. Abordagem de instrumentação e monitoramento SRE. Application Performance Management (APM). Definição de Dashboard. Monitoramento de aplicações: definição e geração de alertas e relatórios de performance. Utilização de logs, métricas e tracing. Métricas e medição de maturidade para DevOps. OpenTelemetry.

## **Disciplina 5: MODELAGEM E ORGANIZAÇÃO DE DADOS**

Ementa: Requisitos de informação. Vocabulário de dados. Tipos de dados. Padrões para modelagem de dados. Modelagem dos diversos tipos de dados. Desenvolvimento e manutenção de modelos conceituais, lógicos e físicos de dados. Políticas, Procedimentos e Padrões de Modelagem de Dados. Forward and Reverse Engineering. Catálogo de dados e metadados. Processo catalogação de dados. Ferramentas para modelagem de dados, repositório, versionamento de modelos e registros de processos de dados. Estratégias para implementar, desenvolver e testar manutenções em bancos de dados.

## **Disciplina 6: 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 7: 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 8: 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 9: 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 10: 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 11: GESTÃO DE INCIDENTES E RISCOS**

Ementa: Fundamentos de incidentes e riscos. Medidas técnicas e administrativas de prevenção e resposta a incidentes. Lifecycle de Gerenciamento de Incidentes: Processo e Etapas. Técnicas e padrões e práticas para gestão de incidentes. Produzindo e publicando Post-mortems. Tipos de riscos no contexto de tecnologia da informação. Classificação de ativos e serviços. Tolerância ao Risco. Ciclo de vida do gerenciamento de riscos. Metodologias para mensurar riscos. Planos de contingência. Ferramentas de planejamento e Comunicação.

### **Disciplina 12: ARQUITETURA DE GERENCIAMENTO DE DADOS**

Ementa: Conceitos e princípios de arquitetura de dados. Conceitos de Data Mesh. Requisitos arquiteturais. Componentes e características da arquitetura de dados moderna: Camadas de dados. SGBDs Relacionais e NoSQL, Data Warehouse e Data Lake, processo ETL e ELT, soluções em processamento distribuído, barramentos de mensageria de dados. Arquiteturas de Referência Abordagens e estratégias para arquitetura de dados. Tradução de requisitos de negócios em especificações técnicas. Requisitos arquiteturais. Elaboração da arquitetura de dados. Melhores práticas para

arquitetura de dados. Gestão da arquitetura de dados.

### **Disciplina 13: IA GENERATIVA PARA GERENCIAMENTO DE DADOS**

Ementa: Fundamentos de IAs Generativas (GenAI). Plataformas de GenAI. Engenharia de Prompt. GenAI e suas implicações na gestão e análise de dados. Abordagens para adoção de Generative AI em arquitetura de dados. Modernização da arquitetura de dados. Automação de processos e procedimentos de gerenciamento de dados. Criação de APIs para acesso a dados. Identificação de oportunidades de produtos de dados. Conceitos de AI-as-a-Service, AI-as-a-commodity, AI Gateways. Tendências.

### **Disciplina 14: 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.

### **Disciplina 15: GESTÃO E QUALIDADE DE DADOS**

Ementa: Gestão de dados: conceitos, benefícios e desafios. Ciclo de vida do dado. Modelos operacionais de gerenciamento de dados. Qualidade de dados: conceito e dimensões. Qualidade dos metadados. Processo de melhoria de análise e avaliação de qualidade de dados. Definição e implementação de procedimentos operacionais de gestão e qualidade de Dados. Construção de soluções de monitoramento de arquiteturas e qualidade de dados. Indicadores, alarmes e métricas de acompanhamento. Serviços de alertas baseados em cloud. Ferramentas de visualização para monitoramento. Definição e implantação de políticas de backup e recovery de dados. Estudo de casos.

### **Disciplina 16: ARQUITETURA DE SOLUÇÕES EM NUVEM**

Ementa: Aspectos da Computação em Nuvem: conceitos, tipos, utilização e fornecedores. Componentes de infraestrutura em nuvem (regiões, zonas de disponibilidade). Desenho de soluções de IaaS de acordo com cenários de: elasticidade, balanceamento de carga, alta disponibilidade e DevOps. Arquiteturas de Cloud, Multicloud e Híbridas. Cenários multi-cloud e cloud híbrida: Interconexão entre nuvens públicas e nuvens privadas. Principais soluções de SaaS e PaaS: front-end, back-end, banco de dados e serverless. Segurança. Estratégias de migração de aplicações para provedores de computação em nuvem. Avaliação de viabilidade técnica e financeira (FinOps). Governança de Nuvem.