

Armazenamento e Processamento Massivo e Distribuído de Dados

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.

Arquitetura de Gerenciamento de Dados

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.

Arquitetura de Soluções em Nuvem

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.

Arquiteturas e Serviços de Data Lakes e Data Warehousing

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.

Bancos de Dados Relacionais e Não Relacionais

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

Cultura e Práticas Dataops e Mlops

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.

Estratégia de Dados

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.

Gestão de Incidentes e Riscos

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.

Gestão e Qualidade de Dados

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.

Governança de Dados

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.

IA Generativa Para Gerenciamento de Dados

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.

Modelagem e Organização de Dados

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.

Monitoramento e Observabilidade

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.

Preparação, Orquestração e Fluxos de Dados

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.



Especialização em ARQUITETURA DE DADOS

OFERTA 1, TURMA 1, MODALIDADE Assíncrono

EMENTÁRIO DAS DISCIPLINAS

Segurança em Armazenamento, Tratamento e Consumo de Dados

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.