

Arquitetura de Dados em Nuvem

Fundamentos da Arquitetura de Dados em nuvem. Componentes chave de uma arquitetura de dados em nuvem. Plataformas e tecnologias de dados. Projeto e implementação de arquitetura de dados. Segurança e conformidade para Arquiteturas de Dados em nuvem. Políticas de backup, recuperação e orquestração de dados. Orquestradores baseados em cloud. Otimização e monitoramento de Dados na Nuvem. Projetos práticos.

Arquitetura de Solução de Dados

Conceitos e princípios de arquitetura de dados: camadas, componentes, princípios e atributos. Papéis e responsabilidades do arquiteto de dados. Ciclo de vida da arquitetura de dados. Tecnologias para arquitetura de dados: 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 para Processamento de Data Stream

Introdução ao processamento de data stream. Arquitetura para processamento de data stream: latência, escalabilidade, tolerância a falhas. Principais plataformas. Modelos de processamento de data stream: batch vs. streaming, micro-batching, processamento contínuo. Serviços gerenciados em cloud. Soluções serverless e orientadas a eventos. Integração com Data Lakes e Data Warehouses. Pipelines de processamento de data stream. Padrões de design para processamento de data stream: filtragem, agregação, enriquecimento, junção. Governança, segurança e monitoramento de dados. Padrões de arquitetura para processamento de fluxo de dados. Projeto e implementação de arquitetura data stream.

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.

Cultura e Práticas DataOps e MLOps

A cultura DataOps e MLOps. Infraestrutura como Código (IaC). Arquitetura de soluções de Machine Learning. Gestão e ciclo de vida da pipeline

Data Mesh e Arquiteturas Descentralizadas

Fundamentos de Data Mesh e de arquiteturas descentralizadas de dados. Componentes do Data Mesh. Implementação do Data Mesh. Serviços de Consumo de Dados. Arquiteturas descentralizadas alternativas. Governança em arquiteturas descentralizadas. Tecnologias e ferramentas para Data Mesh. Design e implementação de Data Mesh e serviços de consumo. Desafios e boas práticas. Estudo de casos. Projeto prático. Tendências.

Discovery e Analytics

Fundamentos da descoberta de dados. Fundamentos e requisitos de aplicações de suporte a decisão. Princípios de projeto, arquitetura e construção de aplicações OLAP. Análise, visualização e comunicação de dados. Ferramentas de Data Discovery e Self-Service Analytics.

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.

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 Qualidade de Dados

Requisitos de informação. 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. 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. Monitorando as aplicações. Definição de alertas. Relatórios de Performance de Aplicações. 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.

Segurança e Gestão de Dados

Ciclo de vida do dado. Modelos operacionais de gerenciamento de dados. Fundamentos de segurança de dados: confidencialidade, integridade e disponibilidade. Segurança no ciclo de vida dos dados. Controles de segurança de dados. Segurança de aplicações, redes e sistemas de armazenamento de dados. Gestão de dados: conceitos, benefícios e desafios. Repositório, versionamento de modelos de dados e registros de processos de dados. Auditoria e monitoramento de dados. Mitigação de riscos e conformidade. Boas práticas para garantir a confidencialidade,

integridade e disponibilidade dos dados.

Tecnologias de Big Data

Princípios de processamento e de volumes de dados massivos. Características e propriedades de big data. Big Data ecossistema: Hadoop, Spark, Kafka, database NoSQL, plataforma de dado em nuvem. Funções dos diferentes componentes do Big Data: ingestão de dados, armazenamento, processamento, análise, visualização. Modelos NoSQL. Principais SGBD's NoSQL. Técnicas de processamento paralelo e distribuído. Critérios para adotar e projetar aplicações com databases NoSQL e processamento distribuído. Arquitetura de soluções de Big Data: escalabilidade, confiabilidade, segurança. Projeto e implementação de arquiteturas distribuídas de Big Data.