

# Guia Ágil

---

TechGuide - Alura

---

## Administração de Banco de Dados - DBA

---

### Nível 1

#### ☐ Modelagem de dados:

- A Modelagem de Dados é uma área essencial para a compreensão e organização das informações em um sistema ou banco de dados.
- Nesse conjunto de conteúdos, você aprenderá as técnicas e conceitos fundamentais para modelar dados de forma eficiente e estruturada, visando obter informações relevantes para a tomada de decisões.

#### ☐ Bancos de Dados - Fundamentos:

- Os bancos de dados são sistemas de armazenamento de dados que permitem a organização, a manipulação e o acesso eficiente às informações. Existem dois principais tipos de bancos de dados, sendo eles SQL (Structured Query Language) e NoSQL (Not Only SQL).
- O conhecimento dos fundamentos de bancos de dados, tanto SQL quanto NoSQL, é essencial na engenharia de dados. Os bancos de dados são a base para armazenar e manipular os dados utilizados nas análises e nos processos de tomada de decisão. Compreender os diferentes tipos de bancos de dados permite aos engenheiros de dados escolher a solução mais adequada para cada caso, levando em consideração os requisitos de escalabilidade, desempenho, consistência e flexibilidade. Além disso, a

habilidade de escrever consultas SQL e de utilizar as características específicas dos bancos de dados NoSQL é essencial para extrair, transformar e carregar dados de forma eficiente e eficaz.

#### ☐ **Sistemas operacionais:**

- O conhecimento em sistemas operacionais é crucial para uma pessoa administradora de banco de dados (DBA), pois muitas das tarefas críticas associadas ao gerenciamento de banco de dados estão intrinsecamente ligadas ao sistema operacional sob o qual o banco de dados opera. Os DBAs devem entender como os recursos do sistema operacional, como gerenciamento de memória, processamento e segurança, impactam o desempenho, a estabilidade e a segurança do banco de dados. Além disso, habilidades em sistemas operacionais permitem que os DBAs otimizem o sistema de banco de dados para uso eficiente dos recursos do sistema, executem tarefas administrativas como backups e restaurações de forma mais eficaz e implementem medidas de segurança robustas. Esse conhecimento é essencial para garantir que os bancos de dados sejam confiáveis, seguros e capazes de atender às demandas de performance exigidas pelas aplicações empresariais.

#### ☐ **Big Data - Fundamentos:**

- Big Data refere-se a conjuntos de dados extremamente grandes e complexos, que não podem ser facilmente processados ou gerenciados por métodos tradicionais. Esses conjuntos de dados são caracterizados por seu volume massivo, velocidade de geração e variedade de tipos e formatos. Com a explosão da quantidade de dados gerados diariamente, provenientes de diversas fontes como redes sociais, dispositivos IoT e transações comerciais, o Big Data apresenta desafios e oportunidades para extrair insights valiosos e tomar decisões informadas.
- A análise de Big Data envolve o uso de técnicas e tecnologias avançadas, como armazenamento distribuído, processamento em paralelo e aprendizado de máquina, para explorar e transformar esses dados em informações significativas para empresas, organizações e pesquisadores.

#### ☐ **Cloud - Provedores:**

- Conhecer os principais provedores de serviços em nuvem e suas ofertas.
- Comparar as características e serviços oferecidos por diferentes provedores.
- Entender como selecionar o provedor de serviços em nuvem adequado para suas necessidades.
- Explorar as tendências e inovações no mercado de serviços em nuvem.

## Nível 2

### ☐ Banco de dados Oracle:

- Para um DBA especializado em Oracle, é vital dominar conceitos fundamentais de bancos de dados relacionais (RDBMS) e habilidades avançadas em SQL. Compreender a estrutura e o funcionamento dos RDBMS, é crucial para manter a integridade e a performance dos sistemas. Conhecimentos específicos, como manipulação de texto em SQL e o uso de CTEs são essenciais para a otimização de consultas complexas. O DBA deve ser capaz de gerenciar eficientemente o Oracle Database, garantindo operações robustas e inovadoras para a organização.

### ☐ PostgreSQL:

- Para um DBA especializado em PostgreSQL, dominar os conceitos fundamentais dos bancos de dados relacionais (RDBMS) e as funcionalidades avançadas do sistema é essencial. Compreender a estrutura e o funcionamento do PostgreSQL é crucial para garantir a eficiência e a integridade dos dados, além de assegurar uma performance otimizada dos sistemas. Conhecimentos avançados em SQL, como manipulação de texto e o uso de CTEs, são indispensáveis para a execução eficaz de consultas complexas. Além disso, a capacidade de gerenciar eficientemente o PostgreSQL, incluindo o uso de comandos DML e DDL, triggers, transações e desenvolvimento com PL/pgSQL, é vital para manter operações robustas e inovadoras. Isso permite que o DBA contribua significativamente para a estabilidade e a evolução tecnológica da organização.

## ☐ MySQL:

- Para um DBA especializado em MySQL, é fundamental ter um domínio abrangente desse sistema de banco de dados, que é amplamente utilizado devido à sua confiabilidade e versatilidade. Aprender os conceitos básicos e avançados de MySQL permite ao DBA garantir que os dados estejam sempre disponíveis, seguros e otimizados para o melhor desempenho possível. Conhecimentos em consultas SQL, utilização de funções e operadores, e habilidades em manipulação de dados são essenciais para gerenciar eficientemente o banco de dados, suportando aplicações críticas. Além disso, entender como implementar e gerenciar triggers e Common Table Expressions (CTEs) pode melhorar significativamente a eficiência e a manutenção do banco de dados. O domínio de ferramentas modernas, como Docker, para rodar instâncias do MySQL também é crucial para a flexibilidade e escalabilidade dos ambientes de desenvolvimento e produção. Com essa expertise, o DBA pode assegurar uma operação estável e eficiente, contribuindo para a inovação e a continuidade dos negócios.

## ☐ SQL Server:

- O aprofundamento nos conhecimentos de SQL Server é vital para uma pessoa administradora de banco de dados (DBA), visto que este sistema de gerenciamento de banco de dados da Microsoft é amplamente utilizado por corporações para armazenamento de dados, processamento de transações e análise. Dominar o SQL Server equipa o DBA com habilidades essenciais para efetuar a instalação, configuração, tuning e manutenção, garantindo que o banco de dados funcione com eficiência e segurança.

## ☐ MongoDB:

- O conhecimento em MongoDB, um dos principais bancos de dados NoSQL orientado a documentos, é extremamente valioso para uma pessoa administradora de banco de dados (DBA) devido às características únicas que este oferece. Com o crescimento de aplicações que exigem flexibilidade na modelagem de dados e a capacidade de lidar com grandes volumes de dados não estruturados ou semi-estruturados, o MongoDB apresenta-se como uma solução eficaz. Para o DBA, dominar o MongoDB significa

habilidade em implementar esquemas dinâmicos e realizar consultas complexas com eficiência. Além disso, o MongoDB oferece recursos robustos de escalabilidade horizontal, suportando grandes cargas de trabalho distribuídas, o que é essencial para empresas que processam grandes quantidades de dados em tempo real. O conhecimento dessa tecnologia permite que o DBA contribua significativamente para a agilidade e a inovação tecnológica dentro da organização, garantindo que os sistemas de dados sejam tanto escaláveis quanto adaptáveis às necessidades empresariais em constante mudança.

#### **Data Warehouses:**

- Um Data Warehouse é um sistema de armazenamento de dados projetado para análise de negócios, suportando processamento de transações e consultas complexas. Ele permite a consolidação de dados de várias fontes, fornecendo um ambiente unificado para análise de dados. Com recursos de normalização e limpeza de dados, garante a qualidade e a consistência dos dados armazenados. Soluções de Data Warehouse populares incluem Google BigQuery, Amazon Redshift e Snowflake, cada uma com suas próprias características e benefícios únicos.
- Data Warehouses desempenham um papel crucial na engenharia de dados. Eles fornecem uma infraestrutura que facilita a gestão, o processamento e a análise de grandes volumes de dados. Sua capacidade de armazenar e processar grandes quantidades de dados de várias fontes permite análises avançadas e suporte à tomada de decisões, o que é fundamental no atual cenário empresarial orientado por dados. Os engenheiros de dados projetam, constroem e mantêm data warehouses para garantir que os dados estejam acessíveis, seguros e prontos para análise. Além disso, soluções modernas de data warehouse, como o BigQuery, o Amazon Redshift e o Snowflake, oferecem capacidades de escalabilidade e flexibilidade para lidar com o crescente volume, velocidade e variedade de dados, tornando-se ainda mais valiosos para a engenharia de dados.

#### **Data Lakes:**

- Um Data Lake é uma arquitetura de armazenamento de dados que permite o armazenamento de grandes volumes de dados brutos em seu formato

original, provenientes de diversas fontes e em vários formatos, como estruturado, semiestruturado e não estruturado. Diferentemente do Data Warehouse, que exige que os dados sejam limpos e estruturados antes de serem armazenados, o Data Lake permite o armazenamento de dados "como estão", proporcionando flexibilidade para transformar e processar os dados conforme necessário para fins de análise.

- Data Lakes são fundamentais para a engenharia de dados devido à sua capacidade de lidar com a crescente variedade, volume e velocidade de dados gerados no mundo digital atual. Eles fornecem um meio econômico e escalável para armazenar e analisar grandes volumes de dados de diversas fontes, o que é fundamental para análises avançadas, como aprendizado de máquina e análise preditiva. Os engenheiros de dados são responsáveis pela construção e manutenção de Data Lakes, garantindo a ingestão de dados de maneira eficiente, o gerenciamento seguro dos dados e a implementação de processos de governança de dados. Além disso, os Data Lakes suportam a democratização dos dados, permitindo que diferentes partes interessadas, desde cientistas de dados até analistas de negócios, acessem os dados necessários para gerar insights valiosos.

## **Cloud - Fundamentos:**

- Cloud, ou computação em nuvem é a distribuição de serviços de computação pela Internet usando um modelo de preço pago conforme o uso. Uma nuvem é composta de vários recursos de computação, que abrangem desde os próprios computadores (ou instâncias, na terminologia de nuvem) até redes, armazenamento, bancos de dados e o que estiver em torno deles. Ou seja, tudo o que normalmente é necessário para montar o equivalente a uma sala de servidores, ou mesmo um data center completo, estará pronto para ser utilizado, configurado e executado.
- Conhecer a diferença entre IaaS, PaaS e SaaS
- Conhecer os maiores provedores de cloud
- Especializar-se em algum provedor

## Nível 3

### ☐ Administração de banco de dados Oracle:

- Dominar conceitos de administração do Oracle é crucial para um DBA porque garante a eficiência, segurança e disponibilidade dos dados críticos de uma organização. Habilidades em otimização de consultas, gerenciamento de índices e manutenção de transações são essenciais para assegurar que os sistemas de banco de dados funcionem com alta performance e minimizem o tempo de inatividade. Além disso, a capacidade de diagnosticar e resolver problemas rapidamente mantém a integridade e a consistência dos dados, permitindo que a infraestrutura de TI suporte as operações empresariais de maneira robusta e confiável. Essa expertise é fundamental para que o DBA não apenas mantenha os sistemas estáveis, mas também contribua para a melhoria contínua da infraestrutura de dados da empresa.

### ☐ Administração do PostgreSQL:

- Para um DBA, dominar a administração do PostgreSQL é vital para garantir que os sistemas de banco de dados operem com alta eficiência e segurança. O PostgreSQL é conhecido por sua robustez e flexibilidade, mas para tirar o máximo proveito dessas qualidades, é necessário um profundo entendimento de como realizar backups, restaurações, e otimizar o desempenho das consultas SQL. Habilidades em gerenciamento de transações asseguram a integridade e a consistência dos dados, fundamentais em ambientes onde a precisão dos dados é crítica. Além disso, a capacidade de monitorar e ajustar a performance do sistema permite que o DBA mantenha uma infraestrutura ágil e confiável, capaz de suportar as crescentes demandas de aplicações modernas. Esses conhecimentos não apenas mantêm a operação dos bancos de dados estável, mas também promovem melhorias contínuas que impulsionam a eficiência e a inovação dentro da organização.

### ☐ Administração do MySQL:

- Para um DBA, dominar a administração do MySQL é essencial para garantir a eficiência, segurança e disponibilidade dos dados. MySQL é amplamente



utilizado devido à sua confiabilidade e facilidade de uso, mas para maximizar seu potencial, o DBA precisa ser proficiente em áreas como backup e recuperação de dados, otimização de consultas e gerenciamento de índices. Essas habilidades asseguram que os sistemas de banco de dados operem com alto desempenho e que os dados sejam recuperáveis em caso de falhas. Além disso, a compreensão profunda de como manter a integridade e a consistência dos dados através do gerenciamento de transações é crucial em cenários onde a precisão dos dados é vital. Ao dominar essas práticas, o DBA pode manter uma infraestrutura de banco de dados robusta, segura e otimizada, essencial para suportar as operações críticas e o crescimento da organização.

#### ☐ **Administração do SQL Server:**

- Para um DBA especializado em SQL Server, dominar a administração desse sistema de banco de dados é fundamental para garantir a segurança, a integridade e o desempenho das operações de dados da organização. O SQL Server é amplamente utilizado em ambientes empresariais devido à sua robustez e flexibilidade, mas para maximizar seu potencial, o DBA deve ser proficiente em práticas essenciais como backup e restauração, automação de tarefas de manutenção e aplicação de melhores práticas de segurança. Esses conhecimentos são cruciais para proteger os dados contra falhas e ataques, além de assegurar que as consultas SQL sejam executadas de forma otimizada. Habilidades em gerenciamento de transações também são vitais para manter a consistência e a precisão dos dados, especialmente em operações críticas. Com um sólido entendimento desses aspectos, o DBA pode garantir que o SQL Server funcione de maneira eficiente e confiável, suportando as demandas operacionais e estratégicas da organização.

#### ☐ **Administração do MongoDB:**

- Para um DBA especializado em MongoDB, dominar a administração desse banco de dados NoSQL é crucial para garantir a eficiência, a segurança e a escalabilidade dos sistemas.
- O MongoDB é amplamente utilizado em aplicações modernas devido à sua flexibilidade na manipulação de grandes volumes de dados e na modelagem dinâmica. Habilidades em gerenciamento de clusters, segurança de dados,



e técnicas de backup e recuperação são fundamentais para assegurar que os sistemas funcionem sem interrupções e estejam protegidos contra falhas e ataques. Além disso, a capacidade de otimizar consultas e gerenciar armazenamento eficazmente, incluindo o uso de GridFS para arquivos grandes, permite ao DBA manter um desempenho superior do banco de dados, mesmo em ambientes de alto tráfego.

- Com um sólido entendimento desses aspectos, o DBA pode garantir que o MongoDB atenda às necessidades complexas e crescentes da organização, suportando tanto a operação diária quanto a inovação contínua.

#### ☐ **Proteção de dados:**

- A proteção de dados é o processo de proteger informações importantes de forma que garanta a confidencialidade, integridade e a disponibilidade destes dados.

#### ☐ **Governança de Dados:**

- Governança de dados é um conjunto de políticas, processos e práticas que estabelecem a responsabilidade, a integridade, a qualidade e o uso adequado dos dados em uma organização. Ela envolve a definição de regras e padrões para a coleta, armazenamento, gerenciamento, compartilhamento e uso dos dados, garantindo que sejam confiáveis, consistentes e seguros ao longo de sua vida útil.
- A governança de dados desempenha um papel fundamental na engenharia de dados, pois garante que os dados sejam gerenciados de forma eficiente, confiável e em conformidade com as regulamentações e políticas internas. Ela estabelece diretrizes para a gestão dos dados, incluindo a definição de metadados, a identificação de proprietários de dados, a documentação de políticas de acesso e privacidade, a implementação de medidas de segurança e a garantia da qualidade dos dados.
- A governança de dados também promove a colaboração entre as equipes, estabelecendo processos de tomada de decisão baseados em dados confiáveis e padronizados. Isso resulta em uma melhor qualidade das análises, uma base sólida para tomada de decisões estratégicas e uma

maior confiança nos dados utilizados para impulsionar as iniciativas de engenharia de dados.

## Habilidade Auxiliar: Infraestrutura

### ☐ Redes de Computadores - Fundamentos:

- Rede de computadores é uma malha que interliga milhares de sistemas computacionais para a transmissão de dados. Também conhecidos como nós, esses dispositivos interconectados enviam, recebem e trocam tráfego de dados, voz e vídeo, graças ao hardware e software que compõe o ambiente.
- Configurações de redes são essenciais para que seja possível acessar uma aplicação, principalmente se ela estiver na nuvem.
- Entenda melhor os componentes de rede e a suas atribuições
- Diferenciar os serviços disponibilizados a nível de rede
- Saber as diferenças entre as camadas de rede
- Saber o que é servidor web, proxy reverso e load balancer

### ☐ Contêineres:

- Os contêineres são pacotes de software que contêm todos os elementos necessários para serem executados em qualquer ambiente. Gerenciamento de contêineres é uma área crucial na computação em nuvem e DevOps, que envolve o uso de tecnologias para automatizar o processo de criação, implantação, escalonamento e monitoramento de contêineres. Contêineres são unidades de software padronizadas que permitem aos desenvolvedores empacotar todas as dependências de um aplicativo (código, bibliotecas, configurações, etc.) em um único pacote. Isso permite que o aplicativo seja executado de forma consistente em qualquer ambiente de infraestrutura.
- A tecnologia de contêineres, como exemplificada pelo Docker, fornece um ambiente consistente e portátil para desenvolvimento, teste e implantação de aplicativos, o que é vital para o trabalho eficiente de engenharia de dados. Além disso, o Kubernetes, um sistema de orquestração de contêineres, permite o gerenciamento, a automação e a escalabilidade de

aplicações baseadas em contêineres em ambientes de produção. Dominar esses conceitos e tecnologias possibilita a engenheiros de dados construir e manter pipelines de dados eficientes e confiáveis.

- O Kubernetes (também conhecido como k8s ou kube) é uma plataforma de orquestração de containers open source que automatiza grande parte dos processos manuais necessários para implantar, gerenciar e escalar aplicações em containers.
- Isolar seu software para funcionar independentemente
- Implantar software em clusters
- Modularizar seu sistema em pacotes menores
- Conhecer a plataforma Docker
- Conhecer Kubernetes

#### ☐ **Git e GitHub - Fundamentos:**

- Git é um sistema de controle de versão distribuído gratuito e de código aberto projetado para lidar com tudo, desde projetos pequenos a muito grandes com velocidade e eficiência.
- GitHub é um serviço de hospedagem para desenvolvimento de software e controle de versão usando Git.
- Criar um repositório
- Clonar um repositório
- Fazer commit, push e pull de e para o repositório
- Reverter um commit
- Criar branches e pull requests
- Lidar com merge e conflitos

#### ☐ **Python - Fundamentos:**

- Python é uma linguagem de programação de alto nível, de uso geral, amplamente utilizada em aplicações web, desenvolvimento de software, ciência de dados e Machine Learning. Sua filosofia de projeto enfatiza a

legibilidade do código com o uso de indentação significativa. Python é dinamicamente tipada e tem um garbage collector.

- Conhecer os tipos primitivos
- Declarar variáveis, considerando os diferentes tipos
- Usar estruturas condicionais ('if', 'else')
- Conhecer os operadores de atribuição e comparação
- Usar estruturas de repetição e laços ('while', 'for')
- Usar funções, passando parâmetros e argumentos
- Manipular métodos
- Manipular arrays e listas
- Obter dados de uma API
- Criar construtores
- Funções anônimas

## Habilidade Auxiliar: Dados

### ☐ Pipelines e ETL:

- Pipelines e ETL (Extract, Transform, Load) são conceitos essenciais no campo da Engenharia de Dados, desempenhando um papel fundamental na coleta, transformação e integração de dados. Compreender os fundamentos dessas áreas é essencial para aqueles que desejam trabalhar com a gestão e o processamento eficiente de grandes volumes de informações.

### ☐ Business Intelligence (BI) - Fundamentos:

- Business Intelligence é um conjunto de teorias, metodologias, processos e tecnologias que possibilitam a transformação dos dados "crus" em informações extremamente relevantes para tomada de decisão de uma empresa.
- Conhecer o processo de ETL
- Realizar a modelagem e estruturação de tabelas em um Data Warehouse
- Criar visualizações que façam sentido

- Conhecer o PowerBI

## **Extração e Tratamento de Dados:**

- A extração de dados é o processo de coleta ou recuperação de tipos diferentes de dados de uma variedade de fontes, muitos dos quais podem estar mal organizados ou completamente desestruturados.
- Obter os dados que serão analisados
- Tratar os dados obtidos, transformando-os, alterando sua estrutura e valores a fim de deixar a base de dados mais coerente e garantir que os dados que serão trabalhados estejam nas melhores condições para serem analisados
- Utilizar as bibliotecas Pandas e Scikit-learn para tratar os dados