

Arquiteto de Dados – Por Onde Começar em 5 Passos

DADOS — essa palavra de cinco letras está hoje por todos os lados e cada vez mais impulsiona o mundo dos negócios em todos os cantos do planeta.

Não há uma única empresa hoje que não tente (ou pelo menos que não deseje) aproveitar o poder dos dados para encontrar soluções para problemas, tomar decisões informadas, implantar estratégias com eficiência e executar operações de negócios eficazes e econômicas.

De acordo com um artigo da Forbes (link ao final do artigo), 94% das empresas globais disseram estar usando dados e analytics como parte do seu crescimento econômico e digital. E 65% dessas empresas planejam aumentar seus investimentos em análise de dados.

Em meio ao crescimento no uso de dados e a conversa de nível avançado sobre Big Data e Data Science, surgiram várias novas funções organizacionais que estão ajudando as empresas a obter, processar e assimilar dados complexos diariamente, dentro e fora da empresa.

E o Arquiteto de Dados (às vezes chamado Arquiteto de Big Data) é uma função altamente pertinente no mundo orientado a dados de hoje.

Neste artigo vamos trazer dicas valiosas para você que deseja se tornar Arquiteto de Dados.

Boa leitura.

O Que é Arquitetura de Dados?

Arquitetura de Dados é a estrutura organizacional alinhada aos processos de negócios que padronizam o processo de coleta, armazenamento, transformação, distribuição e uso de dados. A estrutura é criada para proteger dados confidenciais, tornando as partes mais relevantes acessíveis por pessoas autorizadas no momento certo.

Em outras palavras, podemos definir Arquitetura de Dados como: “Um conjunto de regras, políticas e modelos que determinam que tipo de dados são coletados e como são usados, processados e armazenados”.

Os princípios da Arquitetura de Dados, nos quais toda a estrutura se baseia, consideram os dados como um ativo definido em parâmetros predefinidos que são acessíveis, compartilháveis, gerenciáveis e podem ser protegidos.

As empresas implementam a Arquitetura de Dados para converter principalmente as necessidades de negócios em requisitos de dados e sistemas, alinhar processos de negócios com sistemas de TI e gerenciar o fluxo complexo de dados e informações dentro da organização.

Isso nos leva à grande questão: quem constrói a Arquitetura de Dados de uma empresa? A Arquitetura de Dados é conceituada e projetada em alinhamento com as necessidades de negócios por um Arquiteto de Dados, uma das funções que mais cresce em importância no universo da Ciência de Dados.

Quem São os Arquitetos de Dados?

Os Arquitetos de Dados são profissionais encarregados de definir políticas, procedimentos, modelos e tecnologias que serão usados para coletar, organizar, armazenar e recuperar informações para a organização.

Um Arquiteto de Dados é um especialista que formula a estratégia de dados organizacionais, incluindo padrões de qualidade de dados, fluxo de dados dentro da organização e segurança de dados. É a visão desse profissional de gerenciamento de dados que converte requisitos de negócios em requisitos técnicos (que então serão executados por um Engenheiro de Dados, que tem um perfil bem técnico).

Como existe um elo crítico entre negócios e tecnologia, a demanda por Arquitetos de Dados qualificados tem aumentado.

Uma pesquisa do Recruiter.com revelou que apenas 3,9% dos profissionais de gerenciamento e arquitetura de dados optaram por trabalhar por conta própria, enquanto impressionantes 96,1% deles foram contratados por organizações que desejavam aproveitar os dados para obter vantagem tática nos negócios.

A mesma pesquisa previu que a demanda por Arquitetos de Dados deve se intensificar à taxa de 15,94% nos próximos anos. O Guia Salarial da Robert Half Technology fixou o salário médio do Arquiteto de Dados em \$ 141.250 dólares americanos (por ano).

Veja abaixo o gráfico de projeção para o crescimento da função de Arquiteto de Dados até 2029 (todos os links de referência estão ao final do artigo).

Se você tem um perfil menos técnico e gosta do trabalho de projetar e planejar soluções, a função de Arquiteto de Dados pode ser uma boa opção, com alta demanda e alta remuneração, no Brasil e no exterior.

Arquiteto de Dados X Engenheiro de Dados

De acordo com o Data Management Body of Knowledge da DAMA International, um Arquiteto de Dados “fornece um vocabulário de negócios comum padrão, expressa requisitos estratégicos, descreve projetos integrados de alto nível para atender a esses requisitos e se alinha com a estratégia corporativa e a arquitetura de negócios de uma empresa”.

Um Arquiteto de Dados entende as necessidades de negócios, explora a estrutura de dados existente e cria um projeto para construir uma estrutura integrada de dados seguros e facilmente acessíveis, alinhados com a estratégia de negócios. O Arquiteto de Dados também define os processos envolvidos no teste e manutenção de bancos de dados e sistemas de armazenamento.

As funções de Arquiteto de Dados e Engenheiro de Dados são comumente confundidas e, pior, usadas de forma intercambiável. Mas o papel de um Arquiteto de Dados é diferente do papel de um Engenheiro de Dados (embora existam algumas similaridades).

Neste debate entre Arquiteto de Dados e Engenheiro de Dados, enquanto o primeiro projeta e planeja a estrutura de dados, o segundo coloca esse projeto em ação para construir a infraestrutura de dados de uma empresa.

Um Arquiteto de Dados leva em conta todas as fontes de dados relativas às operações de negócios e descreve um projeto para integrar, centralizar e manter os dados. Por outro lado, um Engenheiro de Dados é responsável por construir e testar arquiteturas de dados sustentáveis para a organização para facilitar a busca e recuperação de dados. Os Arquitetos de Dados trabalham em estreita colaboração com os Engenheiros de Dados para criar uma arquitetura de dados sólida.

De forma simples e resumida: O Arquiteto de Dados projeta e o Engenheiro de Dados executa.

Quais São as Responsabilidades de Um Arquiteto de Dados?

As funções e responsabilidades de um Arquiteto de Dados incluem:

- Desenvolver e implementar uma estratégia geral de dados que esteja alinhada com os processos de negócios. A estratégia inclui projetos de modelos de dados, padrões de desenvolvimento de banco de dados, implementação e gerenciamento de Data Warehouses e Data Lakes e sistemas de análise de dados, além de pipelines de dados.
- Identificar fontes de dados, internas e externas, e elaborar um plano de gerenciamento de dados alinhado com a estratégia de dados da organização.
- Coordenar e colaborar com equipes multifuncionais, partes interessadas e fornecedores para o bom funcionamento do sistema de dados corporativos.
- Gerenciar a arquitetura de dados end-to-end, desde a seleção da plataforma, desenho da arquitetura técnica e implementação, até finalmente testar e implementar a solução proposta.
- Planejamento e execução de soluções de Big Data utilizando tecnologias de armazenamento e processamento distribuído. Na verdade, as funções e

responsabilidades do Arquiteto de Dados envolvem o gerenciamento completo do ciclo de vida de uma solução de armazenamento.

- Definir e gerenciar o fluxo de dados (pipelines) e disseminação de informações dentro da organização, integrando a funcionalidade técnica, garantindo a acessibilidade, precisão e segurança dos dados.
- Conduzir uma auditoria contínua do desempenho do sistema de gerenciamento de dados, refinar sempre que necessário e relatar imediatamente qualquer violação ou brecha às partes interessadas.

Vamos então discutir as habilidades essenciais necessárias para ter sucesso na função e como se tornar um Arquiteto de Dados.

Por Onde Começar em 5 Passos

Aqui estão os 5 passos que ajudarão você a seguir uma carreira como Arquiteto de Dados:

1- Esteja Familiarizado com LGPD, GDPR e Outras Leis e Padrões Aplicáveis ao Gerenciamento de Dados

Os dados são um ativo corporativo de importância cada vez maior e os governos estão atentos a isso. Leis e padrões que regem os dados surgiram nos últimos anos e seguem evoluindo, com o objetivo de garantir que os dados sejam usados de forma correta e com as devidas permissões de usuários, com especial atenção à privacidade.

A LGPD é a Lei Geral de Proteção aos Dados no Brasil e se você pretende trabalhar como Arquiteto de Dados em território brasileiro conhecer os detalhes dessa Lei é fundamental (é importante para todos os profissionais de dados, mas em especial para o Arquiteto de Dados). A GDPR é a Lei equivalente na Europa e outras Leis e padrões podem ser encontrados em outras regiões do mundo. Onde sua empresa atuar, é sua obrigação conhecer as Leis aplicáveis a dados.

Você não pode simplesmente projetar uma solução de dados que não leve isso em consideração. Por exemplo: Sua empresa recebe cadastro de clientes e depois gostaria de usar esses dados para personalizar ofertas de produtos a esses clientes. Não há nada de errado nisso, mas a LGPD prevê que o cliente deve estar ciente e concordar sobre o que está sendo feito com seus dados. Logo, a empresa deve mencionar isso em seus termos de uso. Se isso não for feito e não houver a concordância dos clientes, os dados não poderão ser usados.

Recomendamos o estudo da LGPD para quem pretende trabalhar como Arquiteto de Dados.

2- Governança de Dados e Metadados, Linhagem de Dados e Data Quality

Um sistema abrangente de Governança de Dados e Metadados usa ferramentas para limpar todos os dados de uma organização, atribuindo a cada item uma pontuação de qualidade.

Esse tipo de sistema sinaliza e reconcilia problemas e organiza dados corrompidos, defeituosos, desatualizados ou redundantes, de acordo com as regras de retenção. Essa é uma das tarefas de um Arquiteto de Dados. Você precisa desenvolver suas habilidades em Governança de Dados conhecendo os processos e ferramentas.

Aqui em nosso Blog temos uma série de 7 artigos sobre Governança de Dados. Acesse o primeiro artigo da série clicando [aqui](#).

A linhagem de dados revela o ciclo de vida dos dados, registrando e mostrando o fluxo completo de dados, do início ao fim. A linhagem de dados é o processo de compreensão, registro e visualização de dados à medida que fluem das fontes de dados para o consumo.

Isso inclui todas as transformações que os dados sofreram ao longo do caminho – como os dados foram transformados, o que mudou e por quê. Isso também é o papel do Arquiteto de Dados. Leia mais sobre esse tema em outro artigo em nosso Blog. Clique [aqui](#).

A qualidade dos dados (Data Quality) é a medida da condição dos dados com base em fatores como precisão, integridade, consistência, confiabilidade e se estão atualizados.

A medição dos níveis de qualidade dos dados pode ajudar as empresas a identificar erros de dados que precisam ser resolvidos e avaliar se os dados em seus sistemas são adequados para atender à finalidade pretendida. Temos também um artigo em nosso Blog sobre o tema. Clique [aqui](#).

Ferramentas e soluções como Stitch, Keboola e Dataedo são boas ferramentas para você desenvolver suas habilidades neste tópico.

3- Domine o Funcionamento de Bancos de Dados Relacionais, Não Relacionais e Data Stores

Os dados são um ativo, certo? Os dados são importantes, certo? Os dados devem residir em algum sistema de armazenamento, certo? Esperamos que você tenha respondido sim para as 3 perguntas anteriores.

Parte do seu trabalho como Arquiteto de Dados ao projetar soluções de dados é definir exatamente onde os dados irão residir durante seu ciclo de vida. Os dados nascem em um banco de dados relacional? Após limpeza e processamento são levados para outro banco de dados relacional? Ou os dados entram na empresa (após extração de redes sociais, por exemplo) e residem em um Data Lake ou Data Store antes de serem limpos e levados para um banco relacional ou não relacional? Todas essas são perguntas que um Arquiteto de Dados deve responder ao projetar a solução de dados.

Desenvolva suas habilidades em bancos de dados e sistemas de armazenamento. Isso é um requisito básico para a função de Arquiteto de Dados. E lembre-se que o sistema de armazenamento pode ser local ou na nuvem, logo Cloud Computing será importante no seu dia a dia.

Ferramentas e soluções como PostgreSQL, MySQL, Oracle, SQL Server, MongoDB, Apache Cassandra, AWS Data Lake, Amazon S3, Azure Data Lake Storage e Databricks são exemplos do que você irá encontrar no mercado de trabalho.

Ahh, e não esqueça de Linguagem SQL. Isso é o mínimo de conhecimento que se espera de um Arquiteto de Dados.

4- Domine o Processo ETL e ELT

ETL é a sigla em inglês para Extração, Transformação e Carga de dados (ELT inverte as operações de carga e transformação). Processos ETL serão os responsáveis por extrair dados de um lado, aplicar transformações necessárias e carregar em um sistema de armazenamento. Fica fácil perceber que processos ETL farão parte de qualquer arquitetura de dados de qualquer empresa. E arquitetura de dados é responsabilidade de quem? Do Arquiteto de Dados.

Logo, dominar como funcionam os conceitos de ETL e ELT, quando e como aplicar os processos e as principais ferramentas usadas será fundamental no dia a dia da função.

Ferramentas e soluções como Pentaho Data Integration, Apache NiFi, Apache Spark, Airbyte, Azure Data Factory, AWS Glue e Amazon Athena estão entre as mais comuns no dia a dia de um Arquiteto de Dados e conhecer pelo menos algumas delas é importante.

5- Domine o Design e a Automação de Pipelines de Dados

Um pipeline de dados é o fluxo de movimentação de dados dentro de uma empresa (processos ETL ou ELT fazem parte desse fluxo). Um pipeline de dados é como se fosse uma linha de produção, por onde os dados vão passando desde a origem e então vão recebendo limpeza, transformação, enriquecimento, documentação, gestão de metadados, governança, tags de identificação, até chegarem no seu destino e serem usados para análise, arquivados ou descartados.

O design e automação de pipelines é parte fundamental do trabalho de um Arquiteto de Dados. Ferramentas e soluções como Apache Airflow, Apache Beam, Dremio, Apache Kafka e Snowflake estão entre as mais comuns no dia a dia de um Arquiteto de Dados e você deve conhecer pelo menos algumas delas.

Embora programação de computadores não faça parte do trabalho de Arquitetos de Dados, muitas das ferramentas citadas acima requerem algum conhecimento em programação. Recomendamos conhecer bem a Linguagem Python. A DSA tem um curso gratuito para ajudar você com isso.

Conclusão

Arquiteto de Dados é uma das funções na área de dados com maior crescimento previsto para os próximos anos e este é o momento ideal para quem deseja atuar em uma função que fornece ainda alta remuneração. Há poucos profissionais capacitados no mercado de trabalho.