

EDGARD JOSEPH KIRIYAMA

POSTECH

DATA ANALYTICS

ANÁLISE DE NEGÓCIOS

# AULA 02

## SUMÁRIO

O QUE VEM POR AÍ? .....	3
HANDS ON .....	4
SAIBA MAIS .....	5
O QUE VOCÊ VIU NESTA AULA? .....	11
REFERÊNCIAS.....	12
PALAVRAS-CHAVE .....	13

EMANDA

## O QUE VEM POR AÍ?

Fala, Data Analytics! Tudo certo?

Nesta disciplina, vamos descobrir como a análise de negócios é uma poderosa ferramenta para tomar decisões estratégicas fundamentadas. Para isso, aprenderemos, nessa aula, sobre a "Ligação com banco de dados". Esse é um tópico essencial para quem trabalha na área de dados, pois navegar pelos intrincados meandros de um banco de dados, extrair informações valiosas e transformá-las em insights de negócios é fundamental para a nossa profissão. Você vai aprender a criar e gerenciar os nossos bancos de dados, utilizando diferentes tecnologias e linguagens de programação para interagir com eles.

Exploraremos o incrível poder do SQL (Structured Query Language). Com as consultas SQL certas, você será capaz de extrair informações específicas do banco de dados com facilidade.

A cereja do bolo será a integração do banco de dados com uma linguagem de programação (como Python). Isso abrirá as portas para a criação de scripts poderosos, capazes de automatizar tarefas, gerar relatórios personalizados e realizar análises sofisticadas. Seus dias de trabalho manual com os dados estão contados!

## HANDS ON

Agora, vamos mergulhar no universo dos Bancos de Dados e explorar o poder da análise de negócios das ferramentas de Data Analytics! Você terá a oportunidade de colocar a teoria em prática e entender melhor sobre a manipulação de dados.

Em seguida, exploraremos o poder do SQL, a linguagem padrão para interagir com bancos de dados relacionais. Você aprenderá a escrever consultas SQL para recuperar informações específicas, realizar cálculos, filtrar dados e muito mais. Será como desvendar os segredos dos dados para obter insights valiosos e impulsionar as estratégias de negócios.

Vamos lá, experts de dados?

Acesse o [GitHub da nossa turma](#) para visualizar os arquivos propostos na disciplina.

## **SAIBA MAIS**

### **INTRODUÇÃO AOS BANCOS DE DADOS**

Os bancos de dados desempenham um papel fundamental nas empresas modernas, fornecendo uma estrutura organizada para armazenar e gerenciar dados essenciais. Segundo Date (2004), um banco de dados é definido como "um conjunto de dados relacionados, representando aspectos do mundo real". Eles são importantes para as empresas pois permitem o armazenamento eficiente, recuperação e análise de grandes volumes de dados, possibilitando a tomada de decisões informadas e estratégicas (Silberschatz et al., 2011).

Existem diferentes tipos de bancos de dados, cada um com suas características e finalidades específicas. Os bancos de dados relacionais, baseados no modelo relacional proposto por Codd (1970), são amplamente utilizados e estruturados em tabelas, com relações definidas entre elas por meio de chaves primárias e estrangeiras. Eles oferecem um ambiente estável e confiável para o armazenamento de dados estruturados, permitindo consultas flexíveis por meio da linguagem SQL (Sistemas de Gerenciamento de Banco de Dados, 2017).

Por outro lado, temos os bancos de dados NoSQL (Não Apenas SQL), que surgiram como uma alternativa aos bancos de dados relacionais, especialmente para lidar com grandes volumes de dados não estruturados ou semi-estruturados. Os bancos de dados NoSQL utilizam diferentes modelos de armazenamento, como chave-valor, documentos, colunas ou grafos, permitindo maior escalabilidade e flexibilidade para lidar com dados não convencionais (Lima et al., 2018).

### **FUNDAMENTOS DE SQL – STRUCTURED QUERY LANGUAGE**

O SQL (Structured Query Language) é uma linguagem de programação utilizada para gerenciar bancos de dados relacionais. Segundo Date (2004), o SQL é uma linguagem padrão que oferece recursos poderosos para a criação, manipulação e consulta de dados armazenados em um banco de dados relacional. Ele permite a definição de estruturas de tabelas, a inserção e atualização de registros, a recuperação de informações por meio de consultas complexas e muito mais.

No estudo de SQL, é essencial compreender a sua sintaxe e os diferentes comandos disponíveis. Date (2004) destaca que a linguagem SQL é dividida em várias categorias, como DDL (Data Definition Language), DML (Data Manipulation Language) e DQL (Data Query Language). Os comandos DDL são utilizados para criar e modificar a estrutura do banco de dados, enquanto os comandos DML são empregados para manipular os dados dentro das tabelas. Já os comandos DQL, como o SELECT, permitem realizar consultas para obter informações específicas do banco de dados.

As consultas SQL desempenham um papel fundamental na análise de dados para fins de Data Analytics. Com a habilidade de extrair informações relevantes de bancos de dados, as consultas SQL permitem que os profissionais de Data Analytics realizem análises avançadas e descubram insights valiosos.

Segundo Groff e Weinberg (2012), as consultas SQL são usadas para recuperar dados específicos de tabelas ou visões em um banco de dados relacional. Através do comando SELECT, é possível definir as colunas desejadas e aplicar critérios de filtragem para obter um subconjunto específico de dados. Essa flexibilidade permite aos analistas de dados selecionar, agrupar e ordenar os dados com base nas necessidades de análise.

Além disso, as consultas SQL também permitem a realização de cálculos e agregações. Por exemplo, a função SUM pode ser usada para calcular a soma de valores em uma coluna, enquanto a função COUNT pode ser aplicada para contar o número de registros em uma tabela. Essas funcionalidades permitem que os profissionais de dados realizem análises estatísticas e obtenham métricas relevantes sobre os dados.

No contexto do Data Analytics, é comum que as consultas SQL sejam combinadas com outras técnicas, como visualização de dados e análise estatística. Após recuperar os dados relevantes por meio das consultas SQL, é possível criar gráficos, dashboards e relatórios para apresentar os resultados de forma clara e concisa.

Quer saber mais sobre SQL para Data Analytics e suas principais funções? No [blog do Medium do Alex Souza](#), ele aplica bem toda essa linguagem SQL para análise de dados.

## **BANCO DE DADOS NOSQL – NOT ONLY SQL**

Os bancos de dados NoSQL (Not Only SQL) surgiram como uma alternativa aos tradicionais bancos de dados relacionais, buscando atender demandas específicas de aplicações modernas, escalabilidade horizontal e flexibilidade no armazenamento e recuperação de dados.

Diferentemente dos bancos de dados relacionais, os bancos de dados NoSQL não utilizam uma estrutura rígida de tabelas com esquemas predefinidos. Em vez disso, eles adotam um modelo de dados flexível, permitindo armazenar dados de forma não estruturada, semiestruturada ou poliestruturada. Isso permite lidar com a variedade crescente de dados gerados pelas aplicações atuais, como documentos, grafos, colunas ou pares chave-valor.

Um dos principais benefícios dos bancos de dados NoSQL é sua capacidade de escalar horizontalmente, distribuindo os dados em vários servidores e permitindo que sejam adicionados mais servidores conforme a demanda aumenta. Essa escalabilidade possibilita lidar com grandes volumes de dados e alta carga de acesso, tornando-os ideais para aplicações web e móveis que exigem respostas rápidas e escalabilidade sob demanda (SADALAGE, 2019).

Outra característica dos bancos de dados NoSQL é sua flexibilidade na modelagem de dados. Como não possuem um esquema fixo, podem ser facilmente adaptados para atender às necessidades específicas de cada aplicação. Isso facilita a evolução e alteração dos dados, sem a necessidade de modificar toda a estrutura do banco de dados.

No entanto, é importante considerar algumas limitações dos bancos de dados NoSQL. Devido à sua natureza flexível, algumas operações complexas, como joins e transações atômicas, podem ser mais difíceis de executar. Além disso, a consistência dos dados pode ser relaxada em favor de maior desempenho e escalabilidade (SADALAGE, 2019).

## **MODELAGEM DE DADOS**

A modelagem de dados desempenha um papel fundamental na área de Data Analytics, pois permite a estruturação e organização dos dados, facilitando sua análise e a extração de insights significativos. Trata-se de um processo que envolve a definição de entidades, atributos e relacionamentos, resultando em uma representação lógica dos dados que reflete a realidade do domínio em questão.

De acordo com Inmon (2005), a modelagem de dados para Data Analytics é uma atividade complexa, que envolve a compreensão dos requisitos de negócios, a identificação dos dados relevantes e a definição das estruturas adequadas para armazenar e relacionar esses dados. É importante destacar que a modelagem de dados pode ser realizada em diferentes níveis, desde o nível conceitual, que envolve a definição de entidades e seus atributos, até o nível físico, que determina como os dados serão armazenados em um banco de dados específico.

Um dos principais modelos de dados utilizados em Data Analytics é o modelo dimensional. Segundo Kimball e Ross (2013), o modelo dimensional é especialmente adequado para análise de dados, pois representa as informações em dimensões que são relevantes para a tomada de decisões. Nesse modelo, os dados são organizados em torno de fatos e dimensões, permitindo que sejam realizadas análises de forma rápida e eficiente.

A modelagem de dados para Data Analytics também deve levar em consideração as técnicas e ferramentas de análise de dados que serão utilizadas. Por exemplo, a utilização de algoritmos de aprendizado de máquina pode exigir a preparação e transformação dos dados de acordo com os requisitos desses algoritmos.

## **SEGURANÇA E PRIVACIDADE DOS DADOS**

A segurança e a privacidade dos dados são preocupações essenciais no contexto de Data Analytics. À medida que as organizações coletam, armazenam e processam grandes volumes de dados, é crucial garantir que essas informações sejam protegidas contra acesso não autorizado e que a privacidade dos indivíduos seja preservada.

Para garantir a segurança dos dados, é necessário adotar medidas de proteção adequadas. Isso inclui a implementação de firewalls, sistemas de detecção e



prevenção de intrusões, autenticação de usuários e criptografia dos dados. É fundamental, também, estabelecer políticas de segurança que governem o acesso aos dados e ações de monitoramento para identificar atividades suspeitas.

Em relação à privacidade dos dados, é importante cumprir as regulamentações vigentes, como o Regulamento Geral de Proteção de Dados (GDPR) na União Europeia e a Lei Geral de Proteção de Dados (LGPD) no Brasil. Essas regulamentações estabelecem diretrizes claras sobre a coleta, armazenamento e uso dos dados pessoais, além de fornecer direitos aos indivíduos em relação ao controle e proteção de suas informações.

Além disso, é essencial implementar práticas de anonimização e pseudonimização dos dados, removendo ou mascarando informações identificáveis para preservar a privacidade dos indivíduos. A aplicação de técnicas de privacidade diferencial, que adicionam ruído estatístico aos dados, também pode ser uma abordagem eficaz, protegendo a identidade dos usuários enquanto ainda permite análises robustas.

## **LEI GERAL DE PROTEÇÃO DE DADOS - LGPD**

A Lei Geral de Proteção de Dados (LGPD) é uma legislação brasileira que busca garantir a proteção e privacidade dos dados pessoais dos cidadãos. Ela estabelece diretrizes e responsabilidades no tratamento de dados por organizações públicas e privadas.

De acordo com a LGPD, os dados pessoais são definidos como "informações relacionadas a pessoa natural identificada ou identificável". Isso inclui qualquer informação que possa identificar direta ou indiretamente um indivíduo, como nome, endereço, CPF, entre outros dados.

A lei traz uma série de princípios que devem ser observados pelas organizações no tratamento de dados pessoais. Entre eles, destaca-se o princípio da finalidade, que estabelece que os dados devem ser coletados e utilizados para propósitos específicos e legítimos, sendo proibido o uso excessivo ou indevido dos mesmos.

Além disso, a LGPD estabelece os direitos dos titulares dos dados, como o direito de acesso, retificação e exclusão de informações pessoais. Os titulares também têm o direito de serem informados sobre como seus dados estão sendo tratados e com quem eles estão sendo compartilhados.

No que diz respeito às sanções, a LGPD prevê penalidades para o descumprimento da lei, como advertências, multas e até mesmo a suspensão das atividades relacionadas ao tratamento de dados.

É importante ressaltar que a implementação da LGPD requer um esforço conjunto das organizações, envolvendo aspectos técnicos, operacionais e jurídicos. É necessário realizar uma análise do fluxo de dados, estabelecer políticas de segurança da informação e designar uma pessoa encarregada pelo tratamento dos dados pessoais.

## O QUE VOCÊ VIU NESTA AULA?

Nessa aula sobre ligações com Banco de Dados, vimos que os bancos de dados são essenciais para armazenar e organizar as informações que serão analisadas. Podemos optar pelos bancos de dados relacionais, com sua estrutura tabelada e uso de chaves, ou pelos bancos de dados NoSQL, que são mais flexíveis e permitem diferentes formatos de dados. A linguagem SQL é a chave para fazer consultas e obter as informações desejadas. Além disso, a segurança e privacidade dos dados são fundamentais, garantindo que as informações estejam protegidas e em conformidade com as regulamentações, como a LGPD.

Com todo esse conhecimento, podemos explorar os dados de forma eficiente e obter insights úteis para impulsionar os negócios. Além disso, trabalhamos na manipulação dos dados, como dar carga de dados em um banco de dados e como manipular os dados e conectar as nossas bases para conseguirmos fazer as nossas análises futuras.

Muito bom, não é verdade? Agora é praticar e entender bem como todo esse conteúdo é aplicado. Bons estudos!

**IMPORTANTE: não esqueça de praticar com o desafio da disciplina, para que você possa aprimorar os seus conhecimentos!**

**Você não está sozinho(a) nesta jornada! Te esperamos no Discord e nas *lives* com os professores especialistas, onde você poderá tirar dúvidas, compartilhar conhecimentos e estabelecer conexões!**

## REFERÊNCIAS

CODD, E. F. **A Relational Model of Data for Large Shared Data Banks**. 1970. Disponível em: <https://www.seas.upenn.edu/~zives/03f/cis550/codd.pdf>. Acesso em: 20 jul. 2023.

DATE, C. **Introdução a sistemas de bancos de dados**. Rio de Janeiro: Ltc, 2023.

INMON, W. **Building the Data Warehouse**. Indianapolis: Wiley, 2005.

KIMBALL, R; ROSS, M. **The Data Warehouse Toolkit: the definitive guide to dimensional modeling**. Indianapolis: Wiley, 2013.

LIMA, M; MACEDO, J; MONTEIRO, M. (2018). **Banco de Dados NoSQL**. São Paulo: Novatec Editora, 2018.

SADALAGE, P; FOWLER, M. **NoSQL Essencial: um guia conciso para o mundo emergente da persistência poliglota**. São Paulo: Novatec, 2019.

SILBERSCHATZ, A; SUNDARSHAN, S; KORTH, H. **Sistema de banco de dados**. Amsterdã: Elsevier, 2011.

## **PALAVRAS-CHAVE**

Análise de Negócios, Análise de Dados, Business Intelligence.

EMENDAS

The background is a dark blue field filled with numerous small, light blue dots. Overlaid on this are several large, flowing, wavy lines in shades of teal, blue, and yellow. These lines create a sense of motion and depth. Scattered throughout the composition are various geometric shapes: a circle containing the number '7' in the upper center, a small circle on the left, a cross-like shape in the lower left, a small circle below it, and a hexagon in the bottom right corner.

POSTECH