

****Potencializando o desempenho com NoSQL****

**Andre Luiz Sazana Waleczki | RM:559685**

**Guilherme Vinícius dos Santos | RM:560564**

**Henrique Caproni Siqueira | RM:560105**

**Renan Thiago Aviz e Silva | RM:560849**

**Thiago Evangelista Dias | RM:559403**

**Versão 3**

**HISTÓRICO DE VERSÕES**

|  |  |  |  |
| --- | --- | --- | --- |
| Versão | Data | Responsável | Descrição |
| 1 | 14/06/2024 | Patrícia Maura Angelini | Versão Inicial Template PBL Fase 5 - CAP 01 - POTENCIALIZANDO O DESEMPENHO COM NOSQL |
| 2 | 18/06/2024 | Rita de Cássia Rodrigues | Revisão acadêmica |
| 3 | 19/03/2025 | Andre Luiz Sazana Waleczki | Criação de conteúdo |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

FICHA CATALOGRÁFICA   
**[NÃO PREENCHER - PARA USO DO DEPTO DE EAD E BIBLIOTECA]**

A000a Sobrenome, Nome

Título [livro eletrônico] / Nome Sobrenome. -- São Paulo : Fiap, 2016.

x MB ; ePUB

Bibliografia.

ISBN 000-00-00000-00-0

Categoria. 2. Subcategoria. S., Nome. II. Título.

CDU 000.000.00

RESUMO

Template para atividade de PBL fase 5 1º ano TSC.

Palavras-chave: PBL. FASE 5. TEMPLATE

LISTA DE Figuras

[Texto

Descrição gerada automaticamente Figura 1 – Algoritmo construído em python com as bibliotecas pandas, seaborn, matplotlib, numpy, math e scipy 14](#_Toc193296977)

LISTA DE QUADROS

[Quadro 1 – Quadro resumo das tarefas do PBL 11](#_Toc169351308)

LISTA DE TABELAS

No table of figures entries found.LISTA DE CÓDIGOS-FONTE

No table of figures entries found.LISTA DE ComandoS de prompt do sistema operacional

No table of figures entries found.

Sumário

[Potencializando o desempenho com NoSQL 11](#_Toc193296710)

[1.Prova de conceito de banco de dados NoSQL 11](#_Toc193296711)

[1.1 Análise de cenários 11](#_Toc193296712)

[1.2 Cenário 1 11](#_Toc193296713)

[1.2.1 Justificativa do cenário 1 11](#_Toc193296714)

[1.2.2 Empresa que usa o cenário 1 12](#_Toc193296715)

[1.3 Cenário 2 12](#_Toc193296716)

[1.3.1 Justificativa do cenário 2 12](#_Toc193296717)

[1.3.2 Empresa que usa o cenário 2 12](#_Toc193296718)

[1.4 Cenário 3 13](#_Toc193296719)

[1.4.1 Justificativa do cenário 3 13](#_Toc193296720)

[1.4.2 Empresa que usa o cenário 3 13](#_Toc193296721)

[2 ANÁLISE DOS DADOS DE VENDAS 13](#_Toc193296722)

[2.1 Quantidade 13](#_Toc193296723)

[2.2 Preço 15](#_Toc193296724)

[2.2 Correlações 16](#_Toc193296725)

[REFERÊNCIAS 18](#_Toc193296726)

[GLOSSÁRIO 18](#_Toc193296727)

# ****Potencializando o desempenho com NoSQL****

## 1.Prova de conceito de banco de dados NoSQL

## 1.1 Análise de cenários

A realização de testes de cenários é essencial para validar a adequação de diferentes bancos de dados NoSQL a necessidades específicas do e-commerce da Melhores Compras. Abaixo, detalhamos os cenários analisados, justificando as escolhas e fornecendo exemplos de empresas que já utilizam as soluções sugeridas.

## 1.2 Cenário 1

Quando um cliente seleciona um produto, a plataforma de e-commerce exibe, adicionalmente, recomendações de outros itens, baseadas nas compras de quem comprou esse produto e em outras promoções correlatas. No contexto atual, esse cálculo está demorando muito tempo para ser feito utilizando estruturas relacionais, dado o volume de dados envolvidos.

## 1.2.1 Justificativa do cenário 1

Para esse cenário, um banco de dados NoSQL do tipo Grafo foi escolhido, pois permite modelar eficientemente relações complexas entre produtos e clientes. A estrutura de grafos possibilita consultas altamente otimizadas, eliminando a necessidade de JOINs e garantindo desempenho superior na recomendação de produtos.

## 1.2.2 Empresa que usa o cenário 1

Empresas como Netflix, Facebook e Amazon utilizam bancos de dados de grafos para recomendações personalizadas. O Neo4j é um dos bancos mais populares para esse tipo de aplicação.

## 1.3 Cenário 2

A definição da entrega de um produto em 24h depende da disponibilidade de estoque do centro de distribuição mais próximo do endereço de entrega. Se o cliente optar por essa entrega rápida, é necessário realizar a reserva no centro de distribuição e atualizar o estoque automaticamente. Nos testes preliminares, o modelo relacional apresentou baixo desempenho devido ao volume de dados e à alta frequência de atualizações.

## 1.3.1 Justificativa do cenário 2

Um banco de dados NoSQL do tipo Colunar foi escolhido para esse cenário, pois é otimizado para leituras e escritas massivas. Além disso, oferece alta escalabilidade e excelente desempenho na busca de informações sobre disponibilidade de estoque, essencial para manter a eficiência da entrega rápida.

## 1.3.2 Empresa que usa o cenário 2

Empresas como Netflix e Twitter utilizam Apache Cassandra, um dos principais bancos de dados colunar. Já o Google Bigtable, utilizado pelo Google Analytics e pelo Snapchat, oferece uma solução gerenciada para escalabilidade extrema.

## 1.4 Cenário 3

A tela de detalhes de um produto recebe constantemente novas informações, como reviews, versões, dados de entrega, imagens e recomendações. Para armazenar esse conjunto dinâmico de informações, um banco relacional tradicional pode ser ineficiente devido à rigidez de seu esquema.

## 1.4.1 Justificativa do cenário 3

Um banco de dados NoSQL do tipo Documento foi escolhido, pois permite armazenar todas as informações do produto em um único documento JSON. Essa estrutura facilita a consulta e atualização dos dados sem a necessidade de JOINs, melhorando significativamente a performance da plataforma.

## 1.4.2 Empresa que usa o cenário 3

Empresas como Amazon e Spotify utilizam bancos de dados NoSQL do tipo Documento, como o MongoDB e o Amazon DynamoDB, para armazenar grandes volumes de dados não estruturados e permitir acessos rápidos e escaláveis.

# 2 ANÁLISE DOS DADOS DE VENDAS

## 2.1 Quantidade

O método do intervalo interquartil (IQR) foi utilizado para identificar outliers na coluna quantidade. Onde foram detectados 11.283 outliers.

A média das vendas foi recalculada sem os outliers, resultando em uma média mais precisa (reduzindo de 568,16 para 319,91).

Uma estimativa de variabilidade foi calculada ignorando os outliers:

Desvio padrão das vendas **sem outliers**: 19.789,14.

Coeficiente de variação: 6.185,84%.

Amplitude interquartil (IQR) das vendas sem outliers: 62,00.

Texto

Descrição gerada automaticamente Figura 1 – Algoritmo construído em python com as bibliotecas pandas, seaborn, matplotlib, numpy, math e scipy

Fonte: Elaborado pela equipe (2025)

Texto

Descrição gerada automaticamente

Figura 2 – Algoritmo construído em python com as bibliotecas pandas, seaborn, matplotlib, numpy, math e scipy

Fonte: Elaborado pela equipe (2025)

## 2.2 Preço

A média geral dos preços foi calculada, e um teste t de amostra única foi aplicado para comparar as médias por região em relação à média da população.

Para todas as regiões (Centro-Oeste, Nordeste, Norte, Sudeste, Sul), os p-valores foram maiores que 0.05, indicando que não há diferença estatisticamente significativa entre a média de preço de cada região e a média geral.

A mesma análise pode ser aplicada às modalidades de pagamento.

Texto

Descrição gerada automaticamente

Figura 3 – Algoritmo construído em python com as bibliotecas pandas, seaborn, matplotlib, numpy, math e scipy

Fonte: Elaborado pela equipe (2025)

Texto

Descrição gerada automaticamente

Figura 4 – Algoritmo construído em python com as bibliotecas pandas, seaborn, matplotlib, numpy, math e scipy

Fonte: Elaborado pela equipe (2025)

## 2.2 Correlações

Correlação alta entre valor e valor\_comissao (0.94)

Isso sugere que a comissão é fortemente influenciada pelo valor do produto. Quanto mais caro o produto, maior a comissão.

Correlação alta entre valor\_total\_bruto e valor\_comissao (0.90)

Indica que o total bruto de vendas está intimamente ligado à comissão paga.

Correlação moderada entre lucro\_liquido e valor (0.76)

Produtos mais caros tendem a gerar mais lucro líquido, mas essa relação não é perfeita.

Correlação baixa entre quantidade e lucro\_liquido (0.24)

Sugere que vender mais unidades nem sempre se traduz em mais lucro, o que pode indicar variações de margem de lucro entre os produtos.

Correlação praticamente nula entre quantidade e valor\_total\_bruto (-0.00)

Isso pode significar que o total bruto de vendas não depende muito do número de unidades vendidas, mas sim do preço individual dos produtos.

Texto

Descrição gerada automaticamente

Figura 5 – Algoritmo construído em python com as bibliotecas pandas, seaborn, matplotlib, numpy, math e scipy

Fonte: Elaborado pela equipe (2025)

Gráfico

Descrição gerada automaticamente

Figura 6 – Matriz de correlação construído em python com a biblioteca matplotlib

Fonte: Elaborado pela equipe (2025)

# REFERÊNCIAS

SOBRENOME, Nome do autor abreviado. **Título do livro.** Local da edição: Editora, ano.

# GLOSSÁRIO

|  |  |
| --- | --- |
| **Termo** | Explicação. |
| **Termo** | Explicação. |