| Nome | RM |
| --- | --- |
| Alex Seiji Okada Sinzato | RM558634 |
| Gianluca Lazzaris Giudici | RM557256 |
| Juan Marques Spiller | RM555880 |
| Pietro Ramos de Almeida Pacheco | RM557745 |
| Rodrigo Hideki Ishibashi Grajave | RM556689 |

# **Potencializando o desempenho com NoSQL**

# 2 Prova de conceito de banco de dados nosql

## 2.1 Análise de cenários

Esses testes permitem simular diferentes condições de uso, incluindo casos extremos ou inesperados, para avaliar como o sistema responde a cada situação. Isso ajuda a identificar possíveis erros, vulnerabilidades ou áreas que precisam de otimização, antes que o sistema seja implantado em um ambiente de produção.Também, testes de cenários fornecem uma visão prática de como a aplicação se comporta, garantindo que ela atenda aos requisitos funcionais e de desempenho esperados pelos usuários. Realizar esses testes é essencial para minimizar riscos, evitar erros críticos e garantir a qualidade do produto final.

## 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 demora muito tempo para ser feito utilizando estruturas relacionais, dado o volume de dados envolvidos.

Para esse primeiro cenário, a TI propôs o uso de um banco de dados ***NoSQL*** do tipo GRAFO.

### Justificativa do cenário 1

Para o 1° cenário, o tipo de banco de dados NoSQL ideal seria o do tipo ***CHAVE-VALOR***, pois é o tipo de banco que possui alta performance e baixa latência, pois são otimizados para operações simples de leitura e escrita, onde a chave é usada diretamente para acessar o valor associado.

Devido a essa capacidade para acessar os dados quase que instantaneamente, ele pode exibir uma resposta em tempo real, sendo ideal para esse tipo de cenário onde a plataforma precisa de velocidade para realizar o cálculo que gera as recomendações. Um banco de dados do tipo Grafo não seria capaz de ter essa performance, pois são feitos para explorar relações complexas, não sendo eficientes quando a demanda de transações é alta.

Outro motivo é a escalabilidade que o banco de dados do tipo CHAVE-VALOR pode proporcionar, pois são facilmente escaláveis horizontalmente, podendo distribuir os dados de forma eficiente em vários nós, ideal para lidar com um grande volume de dados de tráfego de usuários simultaneamente.

**Empresa que usa o cenário 1**

Um exemplo de empresa que utiliza esse tipo de banco de dados NoSQL é a Amazon, que utiliza seu próprio banco de dados intitulado DynamoDB, para suportar seus próprios serviços como a Amazon.com, AWS e Alexa. O DynamoDB é conhecido por sua capacidade de lidar com um grande volume de dados e transações, além de sua escalabilidade, que é uma característica do tipo de banco de dados NoSQL CHAVE-VALOR.

## 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 fast*, é necessário realizar a reserva no centro de distribuição e, automaticamente, atualizar o estoque para atender a outros clientes. Nos testes preliminares com o uso do modelo relacional, o desempenho foi frustrante, influenciado principalmente pelo volume de dados e frequência de atualização.

Para esse segundo cenário, a TI propôs o uso de um banco de dados NoSQL do tipo COLUNAR.

### Justificativa do cenário 2

Para o cenário 2 é utilizado o banco de dados NoSQL do tipo COLUNAR, no qual é o tipo adequado para esse cenário. Pois o tipo COLUNAR tem como características:

* Bancos de dados NoSQL colunares armazenam dados por colunas em vez de por linhas, o que otimiza o armazenamento e o acesso a grandes volumes de dados, especialmente em operações que acessam apenas algumas colunas. Isso os torna ideais para cargas de trabalho analíticas, como agregações e sumarização de dados. A homogeneidade dos dados em cada coluna permite compressão mais eficiente, reduzindo o espaço de armazenamento e melhorando o desempenho. Além disso, esses bancos de dados são projetados para escalar horizontalmente, distribuindo a carga de trabalho entre vários servidores.

Por se tratar de um grande volume de dados e atualizações constantes, o tipo COLUNAR é perfeito para esse cenário, pois sua característica é a otimização de armazenamentos e acesso a grande volume de dados.

### Empresa que usa o cenário 2

Uma empresa que utiliza o banco de dados NoSQL do tipo colunar é a Netflix, na qual utiliza a Apache Cassandra. Cassandra é conhecido por sua capacidade de escalar horizontalmente, suportar grandes volumes de dados e oferecer alta disponibilidade, o que é crucial para a operação global da Netflix, que lida com trilhões de requisições diárias.

Principais características da Cassandra são:

* Cassandra, utilizado pela Netflix, permite adicionar facilmente servidores para escalar o armazenamento e processamento, essencial para lidar com dados de milhões de usuários em tempo real. Sua arquitetura distribuída oferece redundância e ***failover*** automático, garantindo alta disponibilidade. Além disso, é otimizado para operações de escrita rápidas e consultas de leitura distribuídas, fundamentais para o streaming e personalização em tempo real.

## Cenário 3

A tela de detalhes de um produto sempre recebe novas informações e, hoje em dia, possui informações que podem ser armazenadas juntamente com o produto, tais como: ***reviews*** do produto; suas versões; informações de entrega; imagens; recomendações; dicas, entre outras.

Para esse terceiro cenário, a TI propôs o uso de um banco de dados NoSQL do tipo DOCUMENTO.

### Justificativa do cenário 3

Dentro do 3º cenário, é possível justificar a escolha de um banco de dados NoSQL do tipo documento pela sua flexibilidade em lidar com a variedade de informações associadas à tela de detalhes de um produto, como reviews, versões, imagens e recomendações, sem a necessidade de um ***esquema rígido***, essa flexibilidade permite armazenar dados complexos em um único documento, facilitando a consulta e a gestão das informações de forma eficiente.

Também, a escalabilidade horizontal desses bancos de dados garante alta performance e disponibilidade à medida que o volume de dados cresce, assim elas se tornam ideais para aplicações dinâmicas e em constante evolução, como as encontradas em e-commerces de grande porte.

### Empresa que usa o cenário 3

Um exemplo prático é a eBay que utiliza o MongoDB, um banco de dados NoSQL do tipo documento, para gerenciar seu catálogo de produtos.

# 3 ANÁLISE DOS DADOS DE VENDAS

## 3.2 Análise dos dados na perspectiva da coluna quantidade. Existem outliers nos dados disponibilizados? É possível identificar algo em relação às vendas associadas a estes outliers? Justifique sua resposta. Calcule uma estimativa de variabilidade que ignore o efeito desses outliers.

Sim, existem outliers nos dados disponibilizados, eles são 5,63% das vendas.

Os pedidos que têm mais outliers são na região norte do país na categoria de alimentação.

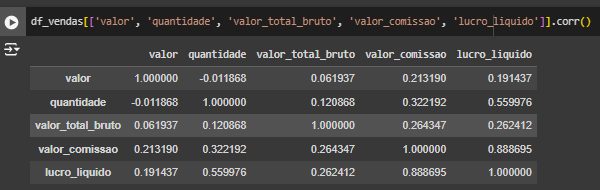
Calculando uma estimativa de variabilidade, podemos fazer o IQR (Intervalo Interquartil) = 1 onde Q3 (terceiro quartil) = 2 e Q1 (primeiro quartil) = 1 a fim de abranger a faixa de valores que contém aproximadamente 50% dos dados centrais.

## 3.3 Em relação à média de preço, há diferença estatisticamente significativa entre a média de preço de alguma região e a média da população? E em relação à média de preço de alguma modalidade de pagamento e à média da população? Justifique a hipótese.

Considerando uma hipótese nula em que a média do valor bruto de pedidos da região sul é igual a média da população e considerando um nível de significância de 5%, não é possível rejeitar a hipótese nula.

Considerando uma hipótese nula em que a média do valor bruto de pedidos da forma de pagamento Boleto Bancário é igual a média da população e considerando um nível de significância de 5%, rejeitamos a hipótese nula.

## 3.4 Calcule a matriz de correlação dos dados fornecidos. Quais as variáveis que apresentam forte correlação positiva ou negativa? Acrescente a matriz de correlação como uma imagem e anexe-a ao seu relatório.

.

# REFERÊNCIAS

AWS. ***O que é um banco de dados chave-valor?***. Amazon. Disponível em: <https://aws.amazon.com/pt/nosql/key-value/>. Acesso em: 17/08/2024

Mongodb. ***Estudos de Caso de Clientes***. Disponível em: <https://www.mongodb.com/solutions/customer-case-studies>. Acesso em: 16/08/2024

AWS. ***Banco de Dados Rápido NoSQL de Chave-Valor – Amazon DynamoDB***. Amazon. Disponível em: <https://aws.amazon.com/dynamodb/>. Acesso em: 17/08/2024

Cassandra apache. ***O que é Apache Cassandra?***. disponível em:

<https://cassandra.apache.org/_/index.html>. Acesso em: 16/08/2024

Livro: NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence

Autores: Pramod J. Sadalage, Martin Fowler

# GLOSSÁRIO

| ***Entrega fast*** | Entrega rápida, geralmente no mesmo dia ou no dia seguinte. |
| --- | --- |
| ***Homogeneidade*** | Retrata similaridade, conformidade, regularidade. |
| ***Failover*** | Redirecionamento automático em caso de falha de sistema para um servidor de backup preparado previamente. |
| ***Chave-Valor*** | Banco de dados que armazena dados em pares chave-valor. |
| ***Esquema Rígido*** | Estrutura fixa de dados em bancos de dados relacionais. |
| ***Reviews*** | Avaliação ou opinião de usuários sobre produtos ou serviços. |
| ***NoSQL*** | Banco de dados não relacional, flexível e escalável. |
| ***e-commerce*** | modelo de comércio baseado na venda de produtos ou serviços em canais digitais. |