Melhores Compras

Com o crescimento da base de usuários e clientes do e-commerce, a base relacional em SQL construída, acaba não atendendo bem o cliente em termos de velocidade, em um mercado competitivo e acelerado, aonde o cliente pode simplesmente ir para outro site ao demorar 10 segundos no carregamento, precisamos garantir que o consumidor finalize suas compras e que o banco de dados não seja um fator de lentidão para o site, isso além de fazer com que as pessoas evitem o Melhores Compras, faz com que nos dias mais cruciais para a companhia, como por exemplo a Black Friday, Dia das Mães etc, o site deixe a desejar impactando diretamente nos resultados do negócio.

Assim sendo, decidimos escolher o tópico **7) Alto consumo de tempo (inviável) para processamento dos dados**. Acreditamos que esse problema seja prioridade para ser resolvido devido a forma como o mercado de vendas online funciona e como isso afeta diretamente o coração do negócio.

De acordo com a pesquisa [A Comparative Study of Relational and Non-Relational Database Models in a Web- Based Application](https://thesai.org/Publications/ViewPaper?Volume=6&Issue=11&Code=ijacsa&SerialNo=11), as FAANG foram as primeiras empresas a descobrir as limitações do modelo relacional, devido as demandas mais novas de uma grande quantidade de dados. Uma grande vantagem do modelo não relacional é a performance e adereçar uma grande quantidade de dados estruturados e não estruturados, contando como uma iteração rápida, code-push frequentes, o estudo foi feito com base em uma aplicação web, assim como o e-commerce, testando assim as diferenças entre o modelo Relacional e Não Relacional.

Após as execuções de SELECT, INSET, UPDATE etc, o resultado da pesquisa conclui que claramente o MongoDB é mais eficiente para um grupo grande de usuários, fazendo assim muito sentido para o Melhores Compras.

Tabela 1. Resultado da Operação INSERT

A table with numbers and letters

Description automatically generated

Tabela 2. Resultado da Operação UPDATE

A table with numbers and letters

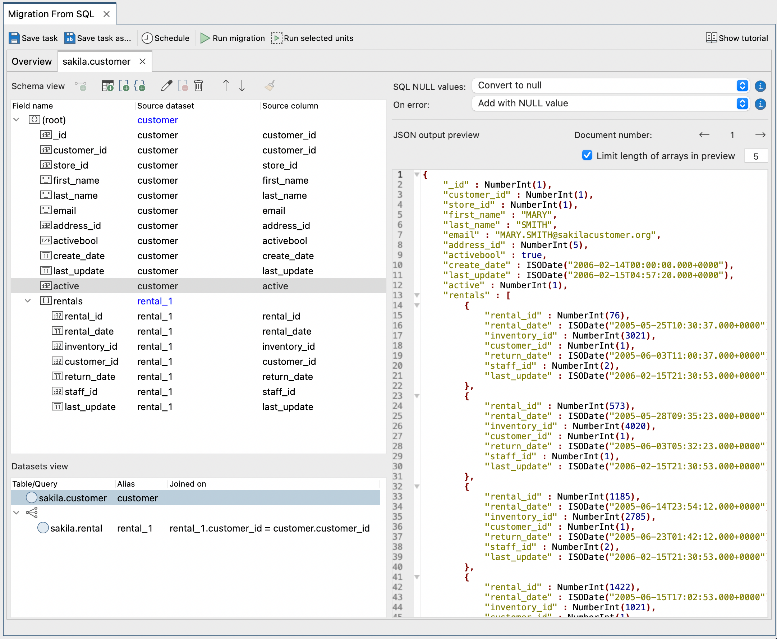
Description automatically generated

Sendo assim, percebemos que o banco de dados NoSQL é escalável horizontalmente e ajudaria na resolução do problema de lentidão, assim como a inserção desse novo modelo de banco de dados em uma arquitetura de Big Data, que entraremos nos detalhes mais a frente.

O primeiro passo seria testar todas as consultas do atual banco de dados, armazenar os tempos de execução para um comparativo posterior, feito isso é preciso replicar o banco de dados inicial porém no MongoDB, isso pode ser feito através do Studio 3T. É importante destacar que o tipo de banco de dados escolhido foi Document Base:

A computer keyboard with many squares

Description automatically generated with medium confidence



Assim vamos realizar a trasição, uma vez estando no MongoDB, podemos iniciar a arquitetura de Big Data, vamos realizar 3 monitoramentos diários, um no próprio banco ,outro no site e por fim diariamente no CRM para monitorar a lentidão e possíveis chamados no CRM de atendimento ao cliente relacionado a lentidão.

1.



2.

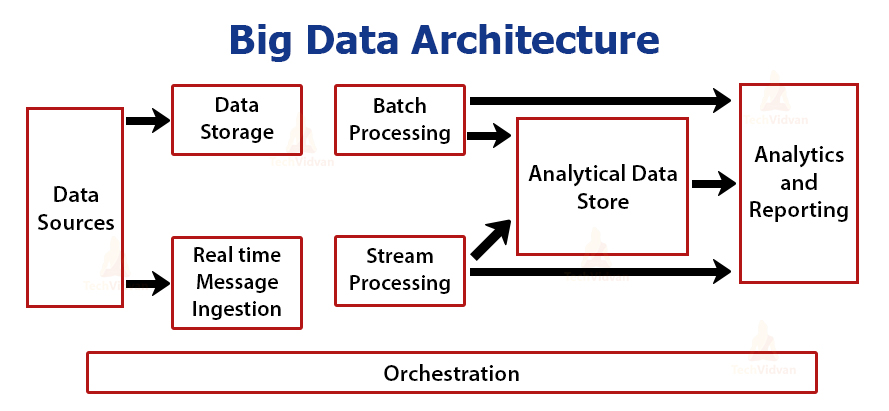


3.



Essas consultas vão nos ajudar a diariamente monitorar o problema de lentidão no banco de dados, que deve já reduzir consideravelmente com a implantação do NoSQL mas agora com a próxima etapa, a implantação da arquitetura de Big Data, a velocidade será ainda maior.

# **Arquitetura Big Data – SGV Melhores Compras**



**Data Source**

Nosso primeiro passo será extrair os dados do banco de dados MongoDB e trazer ao Azure Data Factory, através do Linked Service podemos conectar onosso banco diretamente ao Data Factory, é preciso configurar a conexão, nome do banco de dados e credenciais:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Tendo feito o Linked Service, precisamos especificar o Dataset desejado e as consultas que devem ser feitas, assim que essas consultas de velocidade forem feitas, podemos definir o destino dessa atividade que será o Data Lake.

Um exemplo de consulta que Podemos realizar seria basicamente um SELECT para os vídeos dos produtos, principal função do SGV e também todos os dados desse produto, observando assim o tempo de execução e analisando esses dados, entende o motive dos picos e vales no gráfico de tempo de resposta.

A diagram of a software program

Description automatically generated with medium confidence

Além disso, Podemos também verificar a tabela de chamados, na qual os chamados precisam estar dentro do interval de tempo x para serem considerados normais, desse modo vamos poder localizar os chamados que tem um tempo de criação muito longo.

A green screen with black text

Description automatically generated

Finzalizando as consultas necessárias, vamos fazer o agendamento do pipeline de ingestão de acordo com a frequência e intervalo desejado.

É importante destacar que não somente os dados das consultas agendadas estarão no Data Lake, todos os dados da nossa base NoSQL estarão disponíveis para as futuras analises no nosso Big Data pipeline.

Além da base de dados, o CRM do time de suporte também será utilizado para a coleta de dados e inserção no Datalake, todos os casos relacionados a lentidão, reportados pelos consumidores no SAC, estarão disponíveis para análise em nosso Sistema Big Data.

A screenshot of a computer

Description automatically generated

Com a ferramenta MongoDB Relational Migrator, Podemos transferir os dados do CRM para o modelo document based que consistem nossos dados, assim podemos incluir esses dados em nosso modelo de Big Data e incluir novos tipos de análise.

A screenshot of a computer program

Description automatically generated

Após essa ingestão dos dados vamos para o armazenamento e tratamento desses dados brutos, ou semiestruturados, o Azure Data Lake Storage é um repositório escalável que suporta dados de diversos formatos, atendendo bem nossas consultas no banco de dados, aplicação web e até mesmo CRM interno. Para inserirmos esses dados no Data Lake, precisamos criar uma Azure Data Factory, realizar as configurações iniciais e o destino final dos dados. Por fim, vale destacar que esses dados serão ingeridos continuamente, utilizando conceitos de data streaming, utilizaremos o Apache Kafka para receber esses dados ao inves de um processamento em lotes.

A diagram of a network

Description automatically generated

A screenshot of a computer

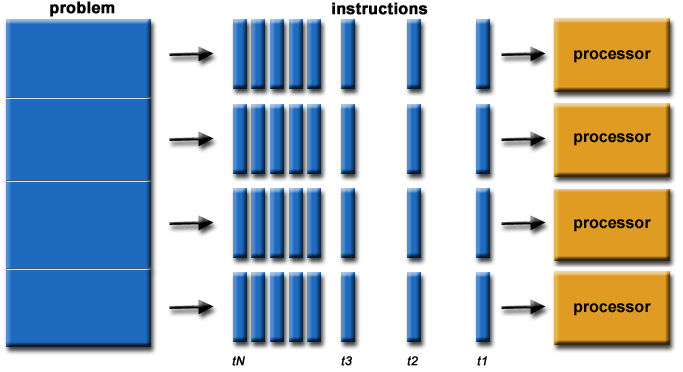
Description automatically generated

**Data Analysis – Compute**

Agora que temos nossos dados inseridos no data lake a partir das consultas agendadas no MongoDB, Podemos iniciar a fase de Data Analysis. Nessa fase, vamos extrair informações relevantes de todas nossas fontes de dados e transformar tudo isso em insights poderosos para a tomada de decisão estratégica, buscando assim eliminar o problema de lentidão entender o padrão dos consumidores impulsionando assim as vendas.

Primeiramente deve ser feito um processo de ETL (Extração, Transformação e Carregamento), ajustando os tipos de dados em um único local, obtendo assim a Integração dos Dados, com esse tratamento e integração, podemos ter uma visão geral do site e cliente.

Após os dados serem tratados, como por exemplo campo de data, tempo em segundos entre outro tratamentos de tipo, podemos iniciar a análise através de algoritmos em python, com a biblioteca numpy para análise de computação numérica, além da SciPy que podemos realizar cálculos estatísticos, para aplicarmos a computação distribuída, trazendo mais eficiência ao processo, podemos usar a bliblioteca PySpark dentro do Apache Spark para realizar os cálculos e análises de forma paralela e distribuída.



Algumas das possíveis análises a serem feitas são:

**Análise de Tempo de Resposta do Site:**

Coletar dados detalhados do tempo de resposta do site, incluindo métricas como tempo de carregamento da página, tempo de resposta do servidor e tempo de renderização.

Utilizar visualizações temporais, como gráficos de linha ou histogramas, para comparar o tempo de resposta em horários de pico com horários normais.

Identificar padrões consistentes de aumento no tempo de resposta durante os horários de pico e correlacionar essas variações com o volume de tráfego.

**Identificação de Padrões de Tráfego:**

Segmentar o tráfego por fontes (por exemplo, orgânico, pago, direto) e tipos de dispositivos (desktop, mobile, tablet).

Analisar o comportamento de navegação dos usuários durante os horários de pico, incluindo páginas mais visitadas e ações realizadas.

Identificar possíveis padrões de tráfego inesperados ou picos repentinos que possam estar causando a lentidão.

**Análise de Carga do Servidor:**

Monitorar a carga do servidor em diferentes componentes, como CPU, memória, e disco.

Identificar os recursos do servidor que estão atingindo limites críticos durante os horários de pico.

Correlacionar picos na carga do servidor com picos de tempo de resposta para entender a relação entre a carga e a lentidão.

**Avaliação da Latência do Banco de Dados:**

Monitorar as consultas ao banco de dados e identificar as mais lentas ou com maior tempo de execução.

Analisar os índices de desempenho das consultas e otimizar consultas lentas através de indexação ou otimização de consulta.

Avaliar a distribuição das consultas por tipo e volume para entender o impacto no desempenho do site.

**Experiência do Usuário:**

Coletar feedback direto dos usuários por meio de pesquisas de satisfação pós-compra ou enquetes sobre a experiência no site.

Analisar as respostas para entender como a lentidão afeta a percepção do usuário e a taxa de conversão.

Identificar padrões nos comentários dos usuários relacionados à lentidão e pontos de melhoria sugeridos.

**Monitoramento de Transações:**

Rastrear o tempo de conclusão das transações, desde a seleção de produtos até o pagamento, e identificar os pontos onde ocorrem atrasos.

Analisar a taxa de conversão durante os horários de pico e correlacioná-la com o tempo de resposta para determinar o impacto da lentidão nas transações.

Identificar se a lentidão afeta a taxa de abandono do carrinho e, se sim, em que etapas do processo.

**Presentation – Use**

Após a análise dos dados armazenados no Data Lake, a apresentação dos resultados no Power BI pode oferecer uma visão clara e interativa das informações, pra isso iremos gerar indicadores realmente relevantes para o negócio e análises contínuas de desempenho, abandon de carrinho e padrões do consumidor.

**1. Taxa de Conversão:**

Gráfico de linha que exibe a taxa de conversão ao longo do tempo.

KPI que mostra a média da taxa de conversão.

**2. Análise de Comportamento do Cliente:**

Segmentação dos clientes por perfil (novos, recorrentes, etc.) usando gráficos de pizza ou de barras empilhadas.

Funil de vendas para visualizar a jornada do cliente no site.

**3. Performance do Produto:**

Gráfico de barras horizontais para mostrar os produtos mais vendidos.

Mapa de calor para representar a distribuição geográfica das vendas dos produtos.

**4. Análise de Tempo de Resposta do Site:**

Medição do tempo de resposta médio do site durante os horários de pico vs. horários normais.

Gráfico de linha para mostrar a variação do tempo de resposta ao longo do tempo.

**5. Categorias de Produtos:**

Gráfico de rosca para representar a distribuição percentual de vendas por categoria de produtos.

**6. Análise de Canais de Marketing:**

Gráfico de barras empilhadas para mostrar o desempenho dos diferentes canais de marketing na geração de vendas.

Gráfico de dispersão para correlacionar o investimento em marketing com as vendas.

**7. Satisfação do Cliente:**

Gráfico de barras ou radar para mostrar a satisfação do cliente com base em pesquisas de feedback.

KPI que exibe a média de satisfação.

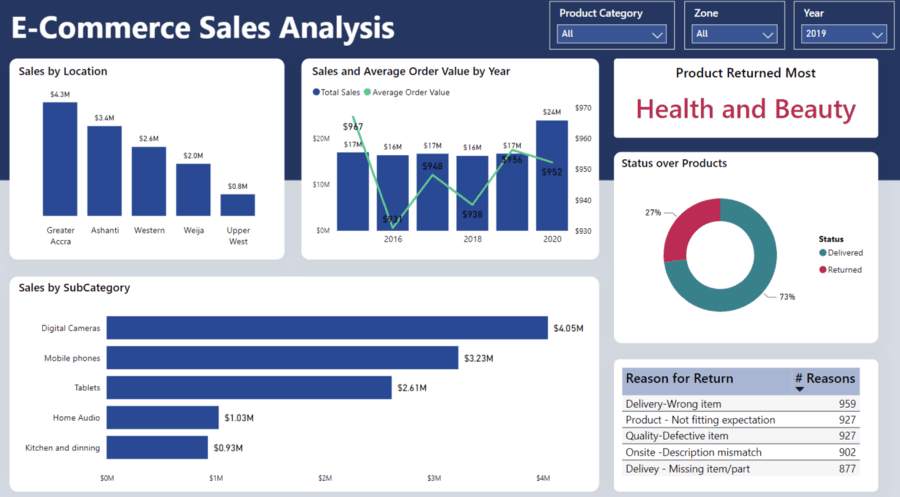
**8. Lentidão do Site:**

Gráfico de área para representar o tempo de resposta do site durante os horários de pico.

Alertas visuais indicando períodos de tempo de resposta anormalmente altos.

**09. Dados Demográficos:**

Gráfico de treemap para exibir a distribuição demográfica dos clientes.



**Big Data Pipeline**

A group of logos on a white background

Description automatically generated

Dscsdd

**Kanban – Project Management**

Para a execução do projeto, foi utilizada a metodologia de Backlog, Doing, Review e Done, conforme mostra a imagem abaixo:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Conclusão**

Concluímos que a adoção de um ambiente Big Data é vital para atender à crescente demanda do mercado de e-commerce, onde a velocidade e eficiência são cruciais. O desafio de garantir que o consumidor finalize suas compras sem enfrentar atrasos significativos devido à performance inadequada do banco de dados é um aspecto central que impacta diretamente o sucesso do negócio, especialmente durante períodos críticos como a Black Friday e o Dia das Mães.

Ao escolher o tópico do alto consumo de tempo para processamento dos dados como prioridade para solução, baseados em estudos comparativos entre modelos de banco de dados, ficou evidente que a transição para o MongoDB, um banco de dados NoSQL, pode oferecer uma solução escalável e eficiente para lidar com grandes volumes de dados estruturados e não estruturados, garantindo uma interação rápida e eficaz com um grande número de usuários.

Além disso, a integração de uma arquitetura Big Data, com a coleta e análise contínua de dados provenientes do banco de dados, do site e do CRM, permite um monitoramento em tempo real da lentidão e uma compreensão mais profunda dos padrões de consumo e comportamento dos usuários. Isso, por sua vez, proporciona insights valiosos para a tomada de decisões estratégicas e aprimoramento constante da experiência do usuário.

O processo envolve a extração, transformação e carregamento dos dados para o Data Lake, seguido por análises detalhadas utilizando algoritmos e ferramentas de análise avançadas. A apresentação dessas análises de forma visual e interativa através do Power BI oferece uma visão abrangente e facilita a identificação de áreas de aprimoramento, impulsionando a satisfação do cliente e, consequentemente, as vendas.

Portanto, o investimento na implementação do ambiente Big Data é essencial para a Melhores Compras, não apenas para resolver os problemas de lentidão, mas também para impulsionar a eficiência operacional e proporcionar uma experiência superior aos clientes, fortalecendo sua posição no mercado competitivo de comércio eletrônico.

Além das soluções tecnológicas e aprofundadas análises de dados, a eficácia da nossa abordagem para otimização da eficiência da Melhores Compras também foi potencializada pela metodologia Kanban. Essa ferramenta, baseada na visualização do fluxo de trabalho, demonstrou ser crucial no gerenciamento ágil e eficiente das atividades do projeto.

A aplicação do Kanban possibilitou uma organização clara das tarefas, desde a sua concepção no Backlog até a conclusão no estágio Done. A equipe pôde priorizar as atividades de maneira adequada, focando nos aspectos mais cruciais para o sucesso do projeto. A visualização do progresso e o acompanhamento em tempo real do estado das tarefas permitiram ajustes imediatos, garantindo uma execução fluida e eficaz.

Assim, fica evidente que o uso estratégico da metodologia Kanban complementou de forma significativa nossa abordagem, consolidando um projeto bem-sucedido e alinhado aos objetivos da Melhores Compras.