

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Gabriel Riquieri Campos

**Preparação de Ambiente *Big Data* para  
Acompanhamento do Usuário em Produtos  
Digitais**

**Uberlândia, Brasil**

**2024**

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Gabriel Riquieri Campos

**Preparação de Ambiente *Big Data* para  
Acompanhamento do Usuário em Produtos Digitais**

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção título de Bacharel em Ciência da Computação.

Orientador: Prof(a) Janaína Maria Bueno

Universidade Federal de Uberlândia – UFU

Faculdade de Computação

Bacharelado em Ciência da Computação

Uberlândia, Brasil

2024

Gabriel Riquieri Campos

## **Preparação de Ambiente *Big Data* para Acompanhamento do Usuário em Produtos Digitais**

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção título de Bacharel em Ciência da Computação.

---

**Prof(a) Janaína Maria Bueno**  
Orientador

---

**Prof. Dr. José Eduardo Ferreira Lopes**

---

**Prof. Dra. Maria Adriana Vidigal de  
Lima**

Uberlândia, Brasil  
2024

# Agradecimentos

Agradeço a minha família por todo o apoio em todos os momentos de minha vida, tenho muita sorte de ter cada um de vocês comigo. Agradeço a todos envolvidos em minhas experiências profissionais pela mentoria e compreensão na situação de equilibrar a vida profissional e a vida acadêmica. Agradeço a minha psicóloga por todo o apoio e ajuda durante a graduação. Agradeço a Prof(a) Janaína Maria Bueno pela orientação didática e gentil. Por fim, agradeço a mim mesmo, por ter acreditado em mim mesmo e nunca ter desistido.

*“Take a gamble that love exists, and do a loving act.” - Red Dead Redemption 2*

---

# Resumo

Este trabalho tem como objetivo investigar e propor uma alternativa genérica, baseada em arquitetura de *Big Data*, para o acompanhamento e análise do comportamento do usuário de produtos digitais no setor de telecomunicações. A monografia foi conduzida por meio de um estudo de caso exemplificativo, com foco na análise de dados de um produto de vendas automáticas e na transição para uma cultura orientada a dados. Utilizando ferramentas da Amazon Web Services (AWS), como S3, Lambda e Step Functions, foi desenvolvida uma estrutura genérica, robusta e escalável capaz de processar grandes volumes de dados gerados por interações de usuários com um *chatbot* de vendas via WhatsApp. A alternativa proposta permitiu o desenvolvimento de um *dashboard* para o monitoramento do fluxo de vendas e com isso a identificação de gargalos e otimização da experiência do usuário. Os resultados mostram que a adoção de uma arquitetura de *Big Data* e uma cultura orientada a dados são fundamentais para aumentar a eficiência operacional, melhorar a tomada de decisão e responder a questão: "Por qual motivo o usuário não se torna um consumidor do produto digital?".

**Palavras-chave:** *Big Data*, Vendas Automáticas, Comportamento do Consumidor, Arquitetura em Nuvem.

# Lista de ilustrações

Figura 1 – Fluxo de <i>Tags</i>	
<i>Fonte - Própria</i> . . . . .	24
Figura 2 – Exemplo de Particionamento no S3	
<i>Fonte - Própria</i> . . . . .	26
Figura 3 – Orquestração AWS StepFunction	
<i>Fonte - Própria</i> . . . . .	29
Figura 4 – Funil de vendas separado por macro etapas	
<i>Fonte - Própria</i> . . . . .	29

# Sumário

1	<b>INTRODUÇÃO</b>	8
2	<b>REVISÃO BIBLIOGRÁFICA</b>	11
2.1	Comportamento do Consumidor e Comportamento do Usuário	11
2.2	Experiência do Usuário - <i>User Experience (UX)</i>	13
2.3	Computação em Nuvem	15
2.4	Processo de Extração, Transformação e Carregamento de Dados	16
3	<b>MÉTODO</b>	19
4	<b>DESENVOLVIMENTO</b>	21
4.1	Vendas Automáticas: Benefícios e Desafios	21
4.2	Transição para uma Cultura Orientada a dados e Reestruturação Tecnológica	23
4.2.1	Implementação de <i>Tags</i> e Monitoramento do Fluxo de Vendas	23
4.2.2	Arquitetura de <i>Big Data</i> em Nuvem	24
5	<b>RESULTADOS</b>	30
6	<b>CONCLUSÃO</b>	32
	<b>REFERÊNCIAS</b>	34

# 1 Introdução

Entender o comportamento do consumidor, de maneira geral, é algo explorado e importante para o mercado desde os modelos mais clássicos de negócio (MENDES, 2014). Com a evolução digital, o comportamento do consumidor precisa ser compreendido também em função de seu comportamento como usuário de produtos digitais para que sejam tomadas decisões mais assertivas por parte das empresas que têm direcionado sua atenção para a análise de dados, utilizando-os como norte para aprimorar a experiência do usuário e identificar gargalos de desempenho com objetivo de maximizar o potencial de suas soluções. Diante do elevado volume de dados, caracterizado pelo contexto do *big data*, torna-se essencial uma arquitetura robusta, escalável e eficiente capaz de suportar essa quantidade massiva de informações (BEZERRA et al., 2023). Essa arquitetura deve assegurar a eficiente extração, armazenamento, processamento e modelagem dos dados.

A importância de distinguir entre comportamento do usuário e do consumidor reside na compreensão de que todos os consumidores são usuários, mas nem todos os usuários são consumidores (ROZADOS; PIFFER, 2009). Esta distinção é particularmente relevante em telecomunicações, onde o comportamento de navegação, consumo de conteúdo e interação com serviços digitais pode não resultar diretamente em uma transação comercial, mas influencia fortemente a satisfação, lealdade e a percepção de marca. Portanto, esse trabalho de conclusão de curso foca em analisar o comportamento do usuário além das métricas de vendas, o que permite uma abordagem mais completa para melhorar a experiência do usuário e otimizar a jornada do cliente.

A partir deste ponto, será utilizado o termo comportamento do usuário com um entendimento mais amplo, que abrange não apenas questões de usabilidade, mas também elementos como satisfação, lealdade e percepção de marca. Esse foco mais abrangente é importante porque a satisfação do usuário é um indicador da qualidade da experiência vivida, que pode afetar sua percepção geral da marca, mesmo sem a realizar uma conversão (OLIVER, 2014). A lealdade, está ligada ao continuo uso e à recomendação de produtos, impactando na retenção de usuários e clientes (DICK; BASU, 1994). Além disso, a percepção de marca, construída via interações consistentes, pode definir o sucesso de uma estratégia a longo prazo (KELLER, 1993). Assim, este trabalho considera o comportamento do usuário sob essa ótica mais completa, buscando formas de otimizá-lo para melhorar não apenas a usabilidade, mas também o relacionamento global entre usuários e empresas.

Com uma visão fundamentada em dados, profissionais de diversas áreas, como marketing, tecnologia e gestão de empresas, de modo geral, são orientados por meio de

números e indicadores diversos a tomar decisões mais embasadas. A busca por essas soluções tem impulsionado uma análise mais profunda sobre o tema, como observado em monografias defendidas por alunos da FACOM e FAGEN com estudos em uma abordagem tradicional de modelagem em bancos de dados (ASSIS et al., 2022) e a preocupação em assegurar qualidade em produtos digitais (OLIVEIRA, 2023).

Em um ambiente empresarial, questões relacionadas ao desempenho e aos resultados de produtos digitais são rotineiramente abordadas. Compreender esse desempenho, sendo positivo ou negativo, é importante para a tomada de decisões e prioridades na gestão do produto. Esse entendimento é alcançado por meio da análise de dados e permite responder perguntas que são realizadas frequentemente como: "Qual é o número de visitantes em nossa plataforma?", "Em qual etapa do processo estamos perdendo retenção?", "Qual o volume de pedidos?". A capacidade de responder a essas perguntas de maneira fundamentada, respaldada por dados, garante decisões estratégicas e também impulsiona melhorias contínuas.

Trabalhos que relacionam o desempenho de produtos digitais e uma arquitetura robusta de *big data* não foram, até o momento, encontrados no Repositório Institucional da FACOM e FAGEN. Entretanto, trabalhos relacionados ao tema que exploram uma alternativa tradicional de banco de dados relacional (ASSIS et al., 2022) são relevantes e permitem comparar abordagens e seus respectivos resultados. Por outro lado, foram identificados trabalhos que apresentam diferentes soluções para coleta, armazenamento e análise de dados voltados para outras finalidades, como soluções de *cloud computing* na indústria de cosméticos, (BUSTAMANTE, 2022), que comprovam a importância, atualidade e utilidade do emprego de ferramenta de análise de dados para melhoria de processos e tomada de decisão empresarial, aliando conhecimento técnico e de gestão.

Dessa maneira, o objetivo deste trabalho de conclusão de curso é propor uma alternativa genérica para o desenvolvimento de uma arquitetura de trabalho com grandes quantidades de dados de maneira robusta e escalável com foco no comportamento do usuário de produtos digitais. Essa arquitetura visa proporcionar *insights* significativos a profissionais de diversas áreas no contexto de produtos digitais, promovendo incrementos e decisões embasadas em dados sobre o desempenho e resultados desses produtos.

Para tanto, foi utilizado o método do estudo de caso com a finalidade de exemplificar a trajetória de elaboração da proposta, demonstrando as necessidades específicas de um determinado segmento de mercado, sendo realizadas algumas validações da solução proposta em uma empresa do segmento. A proposta foi desenvolvida em um contexto de *big data*, e uma abordagem tecnológica e financeira atrativa foi a utilização de recursos de computação em nuvem. Este ambiente ofereceu um alto grau de flexibilidade, permitindo a criação e exclusão de recursos praticamente em tempo real, possibilitando escalabilidade. Dessa maneira, foram utilizados os serviços da Amazon Web Services (AWS), como

AWS S3, AWS Lambda e AWS Step Functions. Os recursos foram manipulados por meio de scripts em Python, uma escolha baseada na versatilidade dessa linguagem e na abundância de bibliotecas disponíveis para manipulação e análise de dados. A abordagem não apenas otimizou os custos financeiros associados ao processamento de grandes volumes de dados, mas também explorou a eficiência proporcionada pelo ambiente de computação em nuvem.

Com o presente trabalho, pretende-se contribuir tanto para a ciência da computação quanto para o ambiente de negócios, pois explora a implementação de uma arquitetura de *big data* que utiliza computação em nuvem para processar grandes volumes de dados. A proposta apresenta uma solução inovadora que combina flexibilidade, escalabilidade e eficiência financeira, algo crucial em um contexto de rápido crescimento de dados. Diferente de abordagens tradicionais, é proposto o uso coordenado de serviços da AWS e a automação com scripts em Python, otimizando processos de extração, transformação e carga de dados (ETL) e permitindo decisões baseadas em dados, o que oferece um diferencial competitivo. Para a ciência da computação, a solução apresenta avanços em termos de otimização de recursos computacionais e integração de tecnologias *serverless*. Já no ambiente de negócios, a proposta melhora a tomada de decisões estratégicas, com *insights* mais rápidos e baseados em dados concretos, impactando diretamente na eficiência e nos resultados financeiros das empresas.

A monografia está organizado em cinco capítulos. No Capítulo 1, é apresentada a introdução, onde são discutidos os objetivos e a relevância do estudo. O Capítulo 2 é dedicado à revisão bibliográfica, abordando os conceitos-chave de comportamento do consumidor, experiência do usuário, computação em nuvem e o processo de ETL. No Capítulo 3, são descritos os métodos utilizados para a elaboração do estudo de caso. O Capítulo 4 apresenta o desenvolvimento da proposta, incluindo a arquitetura de *big data* em nuvem e os resultados obtidos. Finalmente, no Capítulo 5, são apresentadas as conclusões, com uma reflexão sobre as contribuições do trabalho e sugestões para estudos futuros.

## 2 Revisão Bibliográfica

Neste capítulo, são apresentados os principais conceitos e estudos que fundamentam o trabalho. Inicialmente, são abordados os conceitos de comportamento do consumidor e comportamento do usuário, importantes para compreender como as empresas podem utilizar dados para evoluir produtos e serviços. Em seguida, é apresentada a experiência do usuário, *user experience*, para a criação de soluções que atendam às expectativas dos clientes em ambientes digitais. Além disso, a computação em nuvem é explorada como uma tecnologia essencial para o processamento e armazenamento escalável de grandes volumes de dados. Por fim, é analisado o processo de ETL (extração, transformação e carregamento de dados), que é central para a coleta e análise eficiente de dados. Esses temas estão diretamente relacionados ao objetivo deste trabalho, que é propor uma arquitetura robusta de *big data* para otimizar a tomada de decisões baseadas em dados em ambientes empresariais.

### 2.1 Comportamento do Consumidor e Comportamento do Usuário

O comportamento do consumidor refere-se ao conjunto de atividades e processos diretamente envolvidas na obtenção, consumo e disposição de produtos e serviços, incluindo os processos decisórios que antecedem e seguem essas ações (SOLOMON, 2016). Essa definição abrange uma ampla gama de atividades e processos, tanto internos quanto externos ao indivíduo, que influenciam a decisão de compra e uso de produtos e serviços. Isso inclui, mas não se limita a, aplicativos móveis e websites. Compreender esses comportamentos é crucial para o desenvolvimento de produtos que atendam às necessidades e expectativas dos clientes.

Já o comportamento do usuário, em um contexto clássico e tradicional, não digital, começou com estudos simples de usabilidade envolvendo a observação e análise das interações humanas com produtos e serviços físicos. Antes da era digital, a compreensão do comportamento do usuário dependia fortemente de métodos como observações diretas, entrevistas, questionários e estudos de caso. Esses métodos buscavam entender as atitudes, preferências, necessidades e dificuldades dos usuários durante o uso do produto. Por exemplo, no design de produtos, a ergonomia e a facilidade de uso eram avaliadas por meio de testes de usabilidade físicos, onde os participantes interagiam diretamente com o produto em ambientes controlados. Da mesma forma, no setor de serviços, a satisfação do cliente era frequentemente medida através de *feedback* direto ou grupos focais, permitindo às empresas ajustar suas ofertas de acordo com as expectativas dos consumidores (NORMAN, 2013).

É importante destacar que esses modelos clássicos se baseavam fortemente em amostragem por ser impossível entrevistar todos os perfis de usuários e entender de forma aprofundada a interação entre usuário e produto (NORMAN, 2013). Com o advento da internet, novos modelos de negócios digitais, alto volume de dados disponíveis e das tecnologias analíticas avançadas, ocorreu uma mudança ao se estudar o comportamento do usuário, uma compreensão mais profunda das ações dos usuários e novos desafios que serão discutidos ao longo desse trabalho e que estão relacionadas não somente à usabilidade e funcionalidades de soluções tecnológicas, mas também com o comportamento e decisão de compra e uso, aproximando-se dos estudos sobre o comportamento do consumidor.

Embora esses conceitos sejam distintos, a interação entre o comportamento do consumidor e do usuário é um ponto importante, especialmente em ambientes digitais. Todos os consumidores são, por definição, usuários de algum tipo de interface ou tecnologia durante sua jornada de compra. No entanto, nem todos os usuários se convertem em consumidores, o que torna essencial analisar suas ações e comportamentos para entender onde ocorre essa ruptura.

Essa distinção se reflete diretamente nas estratégias empresariais, que precisam capturar dados sobre o comportamento do usuário para entender melhor a jornada de compra. Com o advento das tecnologias de *Big Data* e o uso de arquiteturas em nuvem, é possível acompanhar em detalhes onde o usuário desistiu de uma compra ou onde houve uma quebra na jornada. Esses dados são essenciais para personalizar a experiência e otimizar a conversão de usuários em consumidores, unificando as análises de ambos os comportamentos.

Avanços na área de análise de comportamento do usuário, especialmente no contexto de *Big Data*, incluem o uso de computação em nuvem para armazenamento, processamento e análise de dados. Além disso, o uso de inteligência artificial (IA) para prever comportamentos futuros, personalizar experiências e identificar padrões de uso pode indicar oportunidades de inovação ou necessidade de manutenção (MADEIRA; NEVES; JESUS, 2020). É importante destacar que para possibilitar o uso da IA e análises mais aprofundadas, é de suma importância uma arquitetura nesse ambiente em nuvem. Essas arquiteturas flexíveis mas ao mesmo tempo robustas servem como base para implementação desses modelos de IA.

Por exemplo, em um sistema de vendas automáticas, o comportamento do usuário pode ser analisado ao observar onde ele interrompe sua jornada no fluxo de compra, algo que pode não estar relacionado diretamente à decisão de consumo, mas sim a fatores como usabilidade, entendimento do produto ou problemas técnicos. Essa análise comportamental permite que as empresas identifiquem gargalos que impactam negativamente a experiência do usuário, o que pode influenciar sua conversão em consumidor. Portanto, o comportamento do usuário tem um papel fundamental na otimização do funil de vendas,

permitindo que estratégias sejam ajustadas com base nas interações e pontos de atrito detectados ao longo da jornada.

Embora o usuário nem sempre seja o comprador final, ele exerce uma grande influência nas decisões de compra. Suas interações iniciais com plataformas digitais moldam a percepção sobre a marca e influenciam a decisão de compra de outros consumidores. Dessa forma, analisar como o usuário navega em um ambiente digital e interage com processos é fundamental para ajustar estratégias de marketing e otimizar a conversão de vendas. A aplicação de dados comportamentais permite que as empresas personalizem suas ofertas e melhorem a experiência do cliente ao longo de toda a jornada de compra, impactando diretamente na fidelização e engajamento da marca.

No contexto de produtos digitais em telecomunicações, setor utilizado como exemplo nesta pesquisa, entender o comportamento do usuário é fundamental para o desenvolvimento de soluções que sejam não apenas tecnicamente viáveis, mas também profundamente alinhadas com as expectativas e necessidades dos usuários. Isso inclui desde identificação de problemas em código de aplicações, *bugs*, até a personalização de conteúdo e recomendações. Portanto, a análise do comportamento do usuário, sustentada por uma arquitetura *Big Data* robusta, é um ponto importante para a inovação e a excelência no serviço em telecomunicações, direcionando esforços de desenvolvimento de produto e estratégias de engajamento de forma eficaz e orientada por dados.

## 2.2 Experiência do Usuário - *User Experience* (UX)

A Experiência do Usuário (UX) pode ser definida como todas as interações que um usuário tem com um produto ou serviço e sua percepção sobre essas interações. O conceito de UX cresceu e se popularizou com o advento de produtos digitais, tendo suas raízes em princípios de design centrado no usuário, que datam de várias décadas atrás. Donald A. Norman, um dos primeiros a formalizar o termo "experiência do usuário", enfatizou a importância de entender as necessidades, limitações e contextos dos usuários ao projetar produtos (NORMAN, 2013). A evolução da UX acompanhou o desenvolvimento tecnológico, expandindo o foco da usabilidade e funcionalidade para incluir aspectos como emoções, eficiência e facilidade de uso em experiências mais complexas (SOLOMON, 2016).

A UX é um ponto de conexão entre os estudos sobre comportamento do consumidor e comportamento do usuário. Embora todos os consumidores, em um contexto digital, sejam usuários, nem todos os usuários se tornam consumidores. No entanto, o usuário desempenha um papel crítico como influenciador da decisão de compra e no engajamento com a marca. Em muitos casos, o usuário é o primeiro ponto de contato com o produto digital, navegando, interagindo e experimentando o serviço antes mesmo que ocorra a con-

versão em uma transação comercial. A jornada do usuário não se limita à compra direta; ela abrange o engajamento contínuo com a marca, a fidelidade e até o compartilhamento de *feedbacks* que podem influenciar outros consumidores (NORMAN, 2013).

A UX está intimamente relacionada ao comportamento do usuário. Compreender como os usuários interagem com produtos digitais, suas motivações, dificuldades e necessidades permite que designers, desenvolvedores e gerentes de projeto criem experiências mais intuitivas e satisfatórias (NORMAN, 2013). No setor de telecomunicações, no qual os usuários interagem com uma ampla gama de serviços digitais em diferentes etapas da jornada, uma UX de qualidade é crucial para se destacar em um mercado altamente competitivo (GARRETT, 2006). A análise do comportamento do usuário nessas jornadas, por meio de dados, oferece ideias valiosas que podem ser usadas para aprimorar a UX. Padrões de uso, por exemplo, podem indicar perda de retenção em uma etapa específica da jornada, sugerindo a necessidade de otimização, ou revelar preferências por funcionalidades que podem ser destacadas para melhorar a satisfação e o desempenho do produto.

Ao considerar o comportamento do usuário, é importante observar que a jornada de contato com produtos digitais envolve uma série de pontos de interação, ou *touchpoints*. Esses pontos podem incluir desde a pesquisa inicial de um produto ou serviço, a navegação por sites ou aplicativos, o uso de comparadores de preços, até a interação com *chatbots* ou o atendimento ao cliente. Cada um desses *touchpoints* oferece uma oportunidade de aprimorar a experiência do usuário, com potencial de transformar usuários em consumidores. Mesmo que o usuário não realize a compra, suas interações moldam a percepção da marca, influenciando a reputação e o engajamento da empresa no mercado. (KARTAJAYA; KOTLER; HOOI, 2019).

A aplicação de um método de dados do comportamento do usuário, orientados pela UX, pode revelar oportunidades para inovações disruptivas em serviços de telecomunicações. Essa aplicação pode ocorrer em vários segmentos, mas é interessante especialmente no varejo, como é tratado nesse trabalho, pelo fato de conseguir capturar o sentimento de diversos indivíduos de perfis diferentes. Isso inclui a personalização de serviços, a otimização de fluxos de usuário para aumentar a eficiência e a redução de barreiras para a adoção de novas tecnologias.

Os avanços na área de UX estão ligados ao desenvolvimento de novas tecnologias e metodologias de pesquisa. O uso de dados e inteligência artificial (IA) para analisar grandes volumes de dados de comportamento do usuário tem permitido uma personalização e adaptação das experiências digitais. Além disso, técnicas como o *design thinking*, uma abordagem que busca resolver problemas complexos por meio de um processo iterativo de empatia, definição, ideação, prototipagem e teste, têm promovido uma abordagem mais centrada no usuário (BROWN, 2009). Métodos ágeis de desenvolvimento de produto também têm permitido ajustes rápidos baseados em *feedbacks* reais dos usuários, o que é

fundamental em um ambiente dinâmico como o digital.

A personalização de serviços tem sido um ponto central no aumento da eficiência em produtos e serviços digitais, especialmente com o avanço de tecnologias como a Inteligência Artificial (IA), *Big Data* e aprendizado de máquina. A personalização permite que empresas ofereçam produtos e serviços sob medida, ajustando as experiências dos usuários de acordo com suas necessidades, preferências e comportamentos. Isso resulta em maior satisfação, engajamento e eficiência operacional, uma vez que as interações se tornam mais precisas e eficientes, evitando desperdícios e melhorando a alocação de recursos (FADER, 2020).

## 2.3 Computação em Nuvem

A computação em nuvem é um paradigma que permite o acesso a diversos serviços e recursos de computação, como servidores, armazenamento e aplicações, através da internet e com gestão mínima do provedor de serviços. Este conceito, que começou a ganhar forma nos anos 2000, revolucionou a maneira como as empresas armazenam seus dados e gerenciam sua infraestrutura de forma geral. A computação em nuvem é caracterizada por sua capacidade de proporcionar elasticidade rápida, amplo acesso à rede e recursos compartilhados. Essas características fundamentais permitiram uma transformação digital em diversos setores, incluindo telecomunicações (MELL; GRANCE et al., 2011).

A flexibilidade é importante em um ambiente de telecomunicações dinâmico, permitindo que as empresas ajustem rapidamente seus recursos de TI conforme a demanda. A segurança, é uma prioridade máxima, devido a quantidade significativa de dados sensíveis gerenciados por essas empresas. Provedores de nuvem investem maciçamente em tecnologias de segurança avançadas, oferecendo um nível de proteção muitas vezes superior ao que seria economicamente viável para uma única empresa em um modelo de infraestrutura local. A escalabilidade, a capacidade de aumentar ou diminuir recursos de forma eficiente, é outro benefício chave, permitindo que as telecomunicações atendam tanto aos picos inesperados de demanda quanto ao crescimento orgânico a longo prazo (MELL; GRANCE et al., 2011). Essas características são fundamentais para sustentar a inovação e a competitividade no setor. É importante destacar que as vantagens previamente citadas favorecem no quesito de custo das empresas. Por exemplo, é mais interessante para a empresa alugar um servidor em nuvem e poder desligá-lo quando quiser do que todo o trabalho, custo e tempo para fazer isso localmente (RITTINGHOUSE; RANSOME, 2017).

Dessa forma, os modelos tradicionais de infraestrutura local apresentam limitações significativas em termos de escalabilidade, flexibilidade e custos de manutenção. A necessidade de prever a capacidade máxima e realizar investimentos altos em *hardware* e *software*, sem mencionar os custos contínuos de manutenção e atualização, contrasta

fortemente com o modelo de pagamento conforme o uso oferecido pela computação em nuvem. Além disso, a gestão de segurança em uma infraestrutura local exige investimentos consideráveis e especialização contínua, desafios que são tratados através da adoção de soluções em nuvem (RITTINGHOUSE; RANSOME, 2017).

Dentro do contexto de produtos digitais em telecomunicações, a computação em nuvem habilita uma série de aplicações, desde a hospedagem de plataformas de serviço ao cliente até a análise avançada de *Big Data* para entender o comportamento do usuário. A capacidade de processar e analisar grandes volumes de dados, sem a necessidade de infraestrutura física própria, permite que as empresas de telecomunicações obtenham *insights* sobre as necessidades e preferências dos usuários, otimizando a UX e introduzindo novos serviços personalizados rapidamente no mercado. (RITTINGHOUSE; RANSOME, 2017).

Em suma, a computação em nuvem é um pilar fundamental na transformação digital do setor de telecomunicações, oferecendo as ferramentas necessárias para inovar, escalar e proteger os serviços digitais de maneira eficaz e econômica (FOSTER; GANNON, 2017).

## 2.4 Processo de Extração, Transformação e Carregamento de Dados

ETL, que significa *extract, transform e load*, é um processo consolidado em bancos de dados e serviços de armazenamento de informações que envolve a extração de dados de diversas fontes, sua transformação conforme necessária (que pode incluir limpeza, enriquecimento, entre outros) e, por fim, a carga desses dados transformados em um sistema de destino (NAMBIAR; MUNDRA, 2022). Historicamente, o ETL tem sido um processo importante na construção de armazéns de dados, permitindo que as organizações consolidem dados de várias fontes para análise e relatórios. Desde seu início, na década de 1970, com a emergência dos primeiros bancos de dados e sistemas de informação, o ETL tem evoluído para acompanhar as crescentes demandas por volume, velocidade e variedade de dados, características definidoras do *Big Data*.

Com a ascensão da computação em nuvem, o ETL também se transformou. Soluções modernas de ETL na nuvem, como "Google Cloud Dataflow", "AWS Glue" e "Azure Data Factory", oferecem capacidades de processamento de dados altamente escaláveis, flexíveis e eficientes. Essas plataformas permitem o processamento de dados em tempo real, *streaming*, e em lote, *batch*, suportando as necessidades dinâmicas de análise de dados em telecomunicações. Além disso, avanços na forma como os dados são armazenados ocorreram, como por exemplo a popularização de *data lakes*.

Um *data lake* é um repositório de armazenamento centralizado que permite armazenar grandes volumes de dados brutos em seu formato original, independentemente da fonte ou estrutura. Diferentemente de um *data warehouse* tradicional, que armazena dados em formatos estruturados e processados, um *data lake* é projetado para armazenar uma vasta quantidade de dados não estruturados, semiestruturados e estruturados. Isso inclui tudo, desde textos e imagens até *logs* e *streams* de dados em tempo real. A principal vantagem de um *data lake* é sua flexibilidade e escalabilidade, permitindo que as organizações armazenem dados a um custo relativamente baixo e os processem conforme necessário, usando ferramentas de análise e processamento de dados avançadas para extrair *insights*.(NAMBIAR; MUNDRA, 2022)

A importância do ETL em um contexto de nuvem é ampla. Segundo os próprios provedores de computação em nuvem como a AWS e Microsoft é possível enumerar alguns pontos importantes. Primeiramente, a nuvem oferece uma elasticidade e escalabilidade que são essenciais para o processamento de grandes volumes de dados gerados no setor de telecomunicações. Em segundo lugar, a nuvem proporciona uma redução significativa nos custos de infraestrutura e operação, comparada aos sistemas de ETL tradicionais em infraestruturas locais. Isso permite que as empresas de telecomunicações invistam mais em análise de dados e inovação. Além disso, a segurança e a conformidade, que são críticas para o setor, são reforçadas pelas robustas capacidades de segurança oferecidas pelos provedores de nuvem.

Comparado aos modelos de infraestrutura local, o ETL na nuvem oferece vantagens significativas. Enquanto os sistemas locais exigem investimentos substanciais em *hardware*, *software* e manutenção, o ETL na nuvem opera com um modelo de custo variável, pagando-se apenas pelo que se usa. Além disso, a capacidade de escalar recursos de forma rápida na nuvem permite que as empresas de telecomunicações respondam rapidamente às mudanças nas demandas de dados, um aspecto crítico em um setor caracterizado por rápidas inovações tecnológicas e padrões de uso voláteis dos consumidores (RITTINGHOUSE; RANSOME, 2017).

O processo de ETL é crucial para entender o comportamento do usuário em telecomunicações, pois permite a consolidação, limpeza e organização de dados de interação do usuário de múltiplas fontes, incluindo dados de uso de aplicativos, registros de mensagens e vendas. Ao transformar esses dados brutos em informações estruturadas e analisáveis, os analistas podem identificar tendências de comportamento, preferências e necessidades dos usuários, informando o desenvolvimento de produtos, a personalização de serviços e a melhoria da experiência do usuário. A análise de dados comportamentais suportada por ETL eficaz pode revelar *insights* para a retenção de clientes, desenvolvimento de novos serviços e otimização de redes (KIMBALL; CASERTA, 2004).

Em suma, o ETL desempenha um papel fundamental na arquitetura *big data*

para análise de dados do comportamento do usuário em telecomunicações, especialmente quando implementado em um contexto de nuvem. Sua capacidade de processar grandes volumes de dados de forma eficiente, flexível e econômica é indispensável para as empresas de telecomunicações que buscam inovar e melhorar continuamente a experiência do usuário (KIMBALL; CASERTA, 2004).

## 3 Método

Para este trabalho, foi adotado o método de estudo de caso para exemplificar a investigação das necessidades e as soluções aplicadas em uma grande empresa de telecomunicações. Um estudo de caso, conforme definido por Yin (2009), é uma investigação empírica que analisa um fenômeno contemporâneo dentro de seu contexto da vida real, especialmente quando as fronteiras entre o fenômeno e o contexto não são evidentes.

A escolha de um estudo de caso é justificada pela sua capacidade de fornecer uma compreensão dos processos complexos e dinâmicos em ambientes reais. Ao focar em uma empresa de telecomunicações, foi possível explorar e exemplificar com maior profundidade os desafios e as soluções específicas enfrentadas pela organização, o que seria difícil de alcançar com métodos quantitativos ou experimentais. Essa abordagem qualitativa, do tipo descritiva, permite uma análise interessante e contextualizada, para entender as particularidades da implementação de uma arquitetura de *big data* em nuvem.

O problema identificado neste estudo foi a necessidade de uma arquitetura robusta e escalável para suportar o grande volume de dados gerados pelas empresas de telecomunicações. Especificamente, empresas de grande porte desse segmento, enfrentam dificuldades na extração, transformação e carga (ETL) de dados, além de desafios na análise e na utilização desses dados para tomar decisões estratégicas.

Para demonstrar uma alternativa de como enfrentar o problema, foi realizado uma análise detalhada dos processos atuais da empresa, identificando pontos de falha e gargalos no processamento de dados. Entrevistas com profissionais da empresa e a revisão de documentação foram fundamentais para mapear o fluxo de dados e entender o cenário no contexto de dados. Além disso, o autor deste trabalho fez parte da equipe de desenvolvimento da solução, com a função principal de engenheiro de dados, executando desde atividades técnicas como codificação e construção de diagramas até atividades que fogem um pouco do escopo da função como a difusão da cultura orientada a dados.

Com base nessa exploração, foi feita uma proposição de uma solução que envolveu a adoção de recursos de computação em nuvem, utilizando serviços da Amazon Web Services (AWS) como AWS S3, AWS Lambda e AWS Step Functions. Scripts em Python foram desenvolvidos para manipulação e análise dos dados, aproveitando a versatilidade dessa linguagem e a abundância de bibliotecas disponíveis como Pyspark e Pandas. Essa abordagem não só otimizou os custos financeiros, como também melhorou a eficiência em tempo do processamento de dados.

A escolha por uma abordagem qualitativa descritiva foi feita para capturar a complexidade e a natureza dinâmica do ambiente de telecomunicações das empresas. Essa

abordagem permitiu uma análise das interações entre os processos de ETL e as necessidades de negócios da empresa. Além disso, a abordagem qualitativa facilitou a identificação de *insights* sobre como a tecnologia em nuvem poderiam ser adaptadas, modeladas e em uma arquitetura para atender as demandas da empresa.

A escolha do setor de telecomunicações como exemplo para o desenvolvimento desse estudo de caso foi baseada em critérios como a alta demanda por processamento e análise de grandes volumes de dados, a necessidade constante de inovação tecnológica e a importância estratégica da personalização de serviços. Tal contexto proporciona um cenário interessante para aplicar e avaliar as soluções tecnológicas propostas, que buscam melhorar a eficiência no tratamento e análise de dados, atendendo às necessidades específicas do setor.

## 4 Desenvolvimento

Para contextualização, levantou-se que a empresa de telecomunicações em questão enfrentava um desafio significativo: a implementação de um modelo de "Vendas Totalmente Automáticas". O modelo tem como objetivo permitir que os usuários realizem pedidos sem a necessidade de interação humana em qualquer etapa do processo. Ele tem o potencial de melhorar a experiência do cliente ao oferecer um atendimento ágil e simplificado, além de reduzir custos operacionais eliminando a necessidade de comissionamento de atendentes e vendedores.

Porem, a empresa enfrentou dificuldades consideráveis na sua execução, especialmente no que se refere ao *chatbot* implementado via WhatsApp. Embora o WhatsApp tenha sido uma escolha estratégica devido ao seu alto volume de vendas feitas por humanos e sua viabilidade técnica para automação, o *chatbot* inicialmente não conseguiu atingir as metas esperadas em termos de volume de vendas.

### 4.1 Vendas Automáticas: Benefícios e Desafios

As vendas automáticas surgem como uma solução interessante em um mercado cada vez mais digitalizado e orientado pela experiência do consumidor. Ao permitir que os clientes façam suas compras a qualquer momento, de forma rápida e simples, esse modelo potencializa a satisfação do cliente e fortalece a fidelização. Além disso, a automação do processo de vendas contribui para a eficiência operacional da empresa, reduzindo a necessidade de uma grande equipe de atendimento ao cliente e minimizando erros humanos que poderiam ocorrer durante o processo de venda e reduzindo custos de comissionamento.

Entretanto, a implementação de vendas automáticas apresenta desafios técnicos e estratégicos. Requer a integração eficaz de tecnologias digitais, como *chatbots*, plataformas de *e-commerce* e sistemas de CRM (*Customer Relationship Management*), para garantir que todos os aspectos do processo de venda sejam automatizados. No caso desta empresa, o foco foi em um *chatbot* implementado via WhatsApp, uma escolha estratégica devido à popularidade e viabilidade técnica dessa plataforma.

Apesar do potencial, o *chatbot* via WhatsApp enfrentou dificuldades consideráveis em seu primeiro ano de operação. Mesmo com investimentos substanciais em marketing e promoções, o número de pedidos processados permaneceu baixo, frustrando as expectativas de retorno sobre o investimento. O desempenho insatisfatório gerou um ciclo de atribuição de culpa entre as diferentes áreas da empresa sem que houvesse uma compreensão clara das causas do problema.

A diretoria foi a primeira a expressar preocupações, focando principalmente no retorno sobre o investimento (ROI) das iniciativas relacionadas ao *chatbot*. A baixa conversão de vendas em relação ao capital investido nas campanhas de marketing e no desenvolvimento do *chatbot* gerou uma pressão sobre a equipe comercial. A diretoria não possuía dados concretos para entender as causas do baixo rendimento, o que levou a uma cobrança generalizada, alavancado mais em percepções e resultados financeiros do que em análises objetivas.

A área comercial, sentiu o impacto dessas cobranças. Como o principal elo entre as metas de vendas e a implementação do *chatbot*, a equipe passou a pressionar a equipe de desenvolvimento do *chatbot*, alegando que os baixos números de vendas eram causados por falhas técnicas na plataforma. A narrativa foi reforçada pela ausência de dados claros que pudessem indicar problemas específicos no fluxo de vendas, levando a um clima de frustração e desconfiança.

Por outro lado, a equipe de desenvolvimento do *chatbot* culpava a equipe de marketing pela falta de sucesso do projeto. Argumentava-se que as campanhas de marketing não estavam atraindo o público alvo ou que as mensagens de venda não estavam sendo comunicadas de forma eficaz. A percepção de que o problema estava fora do controle técnico do *chatbot* desviava o foco da análise de dados e da melhoria contínua, perpetuando um ciclo de justificativas baseadas em suposições.

A equipe de marketing, por sua vez, atribuía os problemas ao próprio produto. Argumentava que o *chatbot* não era suficientemente intuitivo ou que os fluxos de compra eram confusos, dificultando a conversão. Sem acesso a dados detalhados que mostrassem onde os clientes estavam abandonando o processo de compra, as campanhas de marketing continuavam a ser lançadas sem a certeza de que estavam abordando os problemas reais.

O ciclo de atribuição de culpa destacou um problema sério: a ausência de uma cultura orientada por dados na empresa. Nenhuma das áreas envolvidas estava baseando suas ações e decisões em dados concretos. Em vez disso, as discussões eram dominadas por suposições e percepções subjetivas, com cada área tentando se distanciar da responsabilidade. Essa falta de orientação por dados impediu que a empresa identificasse e resolvesse os problemas de forma eficiente.

A ausência de uma infraestrutura robusta para coleta e análise de dados significava que decisões críticas estavam sendo tomadas com base em opiniões e intuições, em vez de evidências e indicadores. Isso não apenas gerava a ineficiência, como também criava um ambiente de trabalho tenso, onde a colaboração entre as áreas era substituída por uma busca por culpados. A cultura organizacional estava mais focada em encontrar quem era responsável pelo fracasso do que em compreender as causas do problema e trabalhar em soluções.

Esse contexto mostrou-se prejudicial para a inovação e a melhoria contínua. Em vez de encarar os desafios do *chatbot* como uma oportunidade para aprender e ajustar estratégias, a empresa ficou presa em um ciclo de reclamações e justificativas, sem olhar para os dados disponíveis que poderiam ter guiado as equipes para uma resolução dos desafios. Foi nesse ambiente que se tornou evidente a necessidade de mudar a abordagem dos desafios, adotando uma cultura orientada por dados.

Para quebrar o ciclo de atribuição de culpa e promover uma colaboração entre as áreas, a empresa precisava de uma infraestrutura tecnológica que permitisse a coleta, armazenamento e análise de dados para tomada de decisão. Somente com essa base seria possível identificar os verdadeiros problemas e implementar as correções necessárias de forma precisa e objetiva.

A mudança para uma cultura orientada a dados não foi apenas introduzir novas tecnologias, mas também de promover uma mudança na mentalidade e cultura das equipes. Isso é um processo muito desafiador. Essa mudança, como detalhado nas seções subsequentes, foi importante para que a empresa pudesse superar os desafios iniciais e transformar o *chatbot* em uma ferramenta eficiente de vendas automáticas, guiada em *insights* fundamentados em dados concretos.

## 4.2 Transição para uma Cultura Orientada a dados e Reestruturação Tecnológica

Reconhecendo a necessidade de uma abordagem mais analítica, a equipe de desenvolvimento decidiu ser uma vanguarda de uma mudança cultural, adotando uma orientação orientada a dados para resolver os problemas enfrentados pelo *chatbot*. A chegada de um engenheiro de dados, que também atuava como analista de dados, foi um ponto de importante para essa transição.

### 4.2.1 Implementação de *Tags* e Monitoramento do Fluxo de Vendas

O primeiro passo para criar uma base sólida de dados foi a implementação de *tags* em cada etapa do fluxo de venda do *chatbot*. Essas *tags* permitiam monitorar o progresso dos usuários ao longo da jornada de compra, registrando exatamente em que momento cada usuário interrompia sua jornada de compra. Como exemplificado na Figura 1, ao selecionar um tipo de serviço, uma *tag* com o serviço escolhido seria ativada. Caso o usuário digitasse seu CEP, uma *tag* “Usuario-Digitou-CEP” seria ativada, permitindo a equipe rastrear cada interação.

Uma decisão de negócios tomada pela empresa foi como tratar os casos de *timeout*. Após o usuário permanecer inativo por uma determinada quantidade de tempo, uma *tag*

indicando a etapa específica em que o abandono aconteceu seria ativa e em seguida o usuário deixaria o fluxo automático e seria redirecionado para um humano para tentativa de retomar a jornada de venda.

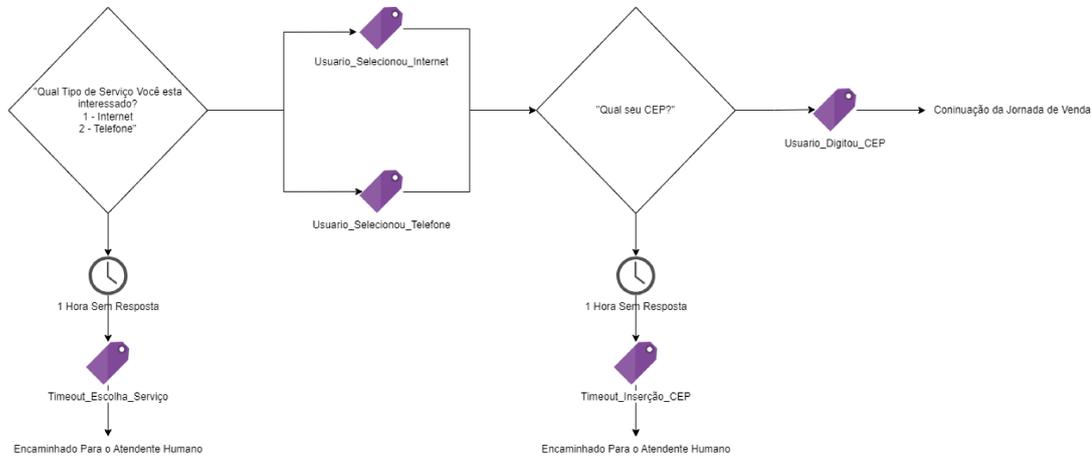


Figura 1 – Fluxo de *Tags*

Fonte - *Própria*

Essa granularidade de dados foi necessária para identificar os *breakpoints* e por sua vez os gargalos e pontos de atrito no fluxo de vendas. Ao analisar essas interações, a equipe poderia entender melhor em que momento na jornada os usuários estavam desistindo do processo de compra e investigar se essas desistências estavam relacionadas a problemas técnicos, questões de usabilidade ou desinteresse nas ofertas apresentadas.

Os dados sobre *tags*, coletados por uma API desenvolvida especificamente para esse propósito, eram armazenados com um identificador único para cada atendimento, juntamente com uma lista das *tags* ativadas durante o processo. No entanto, o grande volume de dados gerado diariamente exigia uma solução tecnológica robusta para garantir que a coleta, o armazenamento e a análise desses dados fossem realizados de forma eficiente.

#### 4.2.2 Arquitetura de *Big Data* em Nuvem

Com a necessidade de processar e analisar um grande volume de dados de forma eficiente, a equipe optou por implementar uma arquitetura de *big data* em nuvem, utilizando serviços da Amazon Web Services (AWS). Essa arquitetura foi projetada para lidar com a complexidade e o volume dos dados gerados pelo *chatbot*, possibilitando que a equipe pudesse extrair informações consistentes.

A arquitetura foi estruturada em torno de um *data lake*, uma solução de armazenamento flexível e escalável que utiliza o serviço de armazenamento S3. Organizada em três camadas lógicas: *raw* (dados brutos), *trusted* (dados confiáveis) e *refined* (dados

refinados). Essa organização permitiu o armazenamento de dados em diferentes níveis de processamento e detalhamento, facilitando consulta e análise.

A camada *raw* é o ponto de entrada de todos os dados no *data lake*. Nessa camada, os dados são armazenados em seu formato original, sem qualquer tipo de processamento ou transformação. Essa abordagem permite a preservação de todos os detalhes dos dados coletados, garantindo que nenhuma informação seja perdida durante o processo de ingestão. Manter os dados em seu formato bruto é interessante para casos de uso futuro, nos quais diferentes tipos de análise podem exigir o acesso às informações originais.

Uma das principais vantagens dessa camada é a flexibilidade que ela oferece. Por manter os dados em seu estado original, a camada *raw* permite que múltiplos processos e transformações possam ser aplicados posteriormente, dependendo das necessidades específicas de análise ou das evoluções nos requisitos de negócios. Além disso, o armazenamento dos dados brutos facilita a conformidade com as políticas de governança de dados, pois garante a rastreabilidade e auditoria das etapas do processamento.

Após a inclusão dos dados na camada *raw*, eles passam por um processo de transformação, limpeza e estruturação antes de serem armazenados na camada *trusted*. Nesta camada, os dados são organizados em um formato tabular estruturado, utilizando o formato Parquet, por conta de sua eficiência em compressão e velocidade de leitura. Esse processo de transformação inclui a validação dos dados, remoção de inconsistências e correções na tipagem dos dados para garantir que as informações estejam consistentes para análise.

A principal vantagem da camada *trusted* é a confiança que ela oferece aos analistas e às ferramentas de BI (*Business Intelligence*) que irão consumir esses dados. Por serem dados que já passaram por um processo de curadoria, diferentes perfis de profissionais podem utilizá-los com a garantia de que estão analisando informações corretas e consistentes. Os dados organizados de maneira estruturada facilitam a execução de consultas e análises, sem a necessidade de realizar grandes transformações adicionais.

A camada *refined* é a última etapa na esteira de dados e é onde os dados são transformados em visões que agregam valor direto ao negócio. Nessa camada, os dados estruturados na camada *trusted* são modelados de acordo com a necessidade para atender a necessidades específicas de análise, como relatórios, *dashboards* ou modelos preditivos. No caso específico deste estudo, a principal visão criada foi um *dashboard* de um funil de vendas, que permitiu a empresa visualizar onde os clientes estavam abandonando o processo de compra.

Uma das maiores vantagens dessa camada é a capacidade de gerar *insights* de forma rápida e eficiente. Como os dados já passaram por todas as etapas de processamento e estão em um formato otimizado, as análises realizadas na camada *refined* podem ser exe-

cutadas com baixa latência, viabilizando respostas rápidas para as perguntas orientadas ao negócio. Essa camada também ajuda na personalização de relatórios e *dashboards*, permitindo que diferentes áreas da empresa acessem as informações que são mais importantes no seu contexto.

Além da organização em camadas, o *data lake* foi particionado por ano, mês e dia, utilizando a estrutura de particionamento do Apache Hive. Esse particionamento tem um impacto significativo na eficiência da arquitetura proposta, especialmente em ambientes de *big data* no qual grandes volumes de dados precisam ser processados. O particionamento dos dados oferece várias vantagens operacionais:

1. **Eficiência:** Particionando os dados, é possível fazer consultas mais rápidas e eficientes, porque o sistema pode ignorar grandes porções de dados que não são relevantes para a consulta em questão. Por exemplo, se um analista quiser consultar dados de um único dia ou mês específico, a consulta será feita apenas para as partições relevantes, reduzindo o tempo de resposta.
2. **Custo:** O particionamento permite que apenas uma fração dos dados seja processada a qualquer momento, reduzindo os recursos computacionais. Isso é importante em ambientes de nuvem, onde o custo de processamento está diretamente relacionado ao volume de dados processados.
3. **Manutenção e Gerenciamento de Dados:** Como exemplificado na *Figura 2*, o particionamento facilita o gerenciamento de grandes volumes de dados, permitindo que a empresa archive ou exclua dados antigos. Por exemplo, dados muito antigos que já não são necessários podem ser movidos para um armazenamento de custo mais baixo ou excluídos.

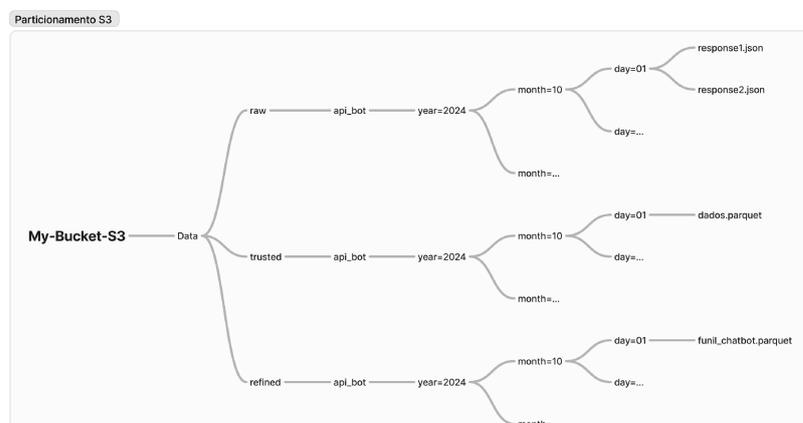


Figura 2 – Exemplo de Particionamento no S3

Fonte - Própria

A combinação das três camadas com a estratégia de particionamento por ano, mês e dia, cria uma arquitetura de *data lake* eficiente e escalável. Essa estrutura permite o trabalho com grandes volumes de dados de maneira eficiente, oferecendo a flexibilidade necessária para se adaptar a diferentes requisitos de negócios.

Essa abordagem melhora a performance geral do sistema e também garante que a escala das operações de dados à medida que o volume e a complexidade dos dados aumentem. A estrutura particionada e a organização em camadas, são pontos da arquitetura que suporta a análise e a otimização contínuas do *chatbot* e de outros escopos também.

Para garantir que os dados fossem coletados, processados e disponibilizados de forma eficiente, foram utilizadas quatro Lambda Functions. Lambda Functions são um serviço de computação sem servidor oferecido pela AWS que permite executar código em resposta a eventos sem a necessidade de gerenciar servidores. Em vez de provisionar, escalar ou monitorar a infraestrutura, as Lambda Functions permitem que os desenvolvedores foquem diretamente na lógica do código, deixando que a AWS cuide da administração da infraestrutura. Esse modelo de computação *serverless* (sem servidor) é útil em cenários onde a carga de trabalho pode variar significativamente ou onde a execução de tarefas é disparada por eventos específicos.

As Lambda Functions escalam automaticamente em resposta ao volume de eventos, possibilitando que o sistema possa lidar com picos de demanda sem a necessidade de provisionar recursos manualmente. Isso é essencial em um ambiente de *big data*, no qual o volume de dados pode variar drasticamente.

Em relação ao custo, as Lambda Functions são cobradas com base no número de solicitações e no tempo de execução, a empresa só paga pelos recursos que efetivamente utiliza. Isso ajuda a otimizar os custos operacionais, especialmente quando comparado com a necessidade de manter servidores dedicados ociosos durante períodos de baixa demanda.

Outro ponto importante é que as Lambda Functions se integram perfeitamente com outros serviços AWS, como S3, DynamoDB, API Gateway, e Step Functions. Essa integração facilita a construção de esteiras de dados complexas, como o que foi necessário para gerenciar o fluxo de dados do *chatbot*.

Para garantir que os dados fossem coletados, processados e disponibilizados de forma eficiente, foi implementada uma arquitetura composta por quatro Lambda Functions, cada uma desempenhando uma função específica dentro do fluxo de dados:

1. Lambda de Extração: Essa função é responsável por realizar as requisições à API que fornece os dados do *chatbot* e armazená-los em formato JSON na camada raw do *data lake*. O uso de Lambda para essa tarefa permite que a função seja acionada automaticamente sempre que necessário, garantindo que os dados sejam coletados de forma contínua e sem falhas. A função particiona os dados por ano, mês e dia,

utilizando a estrutura do Apache Hive, o que otimiza o tempo de consulta e processamento futuro.

2. Lambda de Orquestração: Esta função supervisiona o processo de extração de dados, garantindo que todas as datas relevantes sejam cobertas e que nenhum dado seja perdido. Ela verifica se há dias faltantes no armazenamento de dados e, se necessário, aciona novamente a Lambda de Extração para preencher essas lacunas. Essa função também ajuda a gerenciar eventuais erros, garantindo que o fluxo de dados seja robusto e resiliente a falhas temporárias.
3. Lambda de Transformação: Após os dados serem armazenados na camada raw, essa função transforma os dados de JSON para um formato tabular estruturado, armazenando-os na camada trusted em formato Parquet. Essa transformação é crucial para que os dados possam ser facilmente consultados e analisados posteriormente. O uso do formato Parquet reduz o espaço de armazenamento e melhora o desempenho das consultas.
4. Lambda Refined: Essa função é responsável por criar visões refinadas dos dados, especificamente para atender às necessidades de negócios. No caso do *chatbot*, essa função gera uma tabela que é consumida por um *dashboard* de funil de vendas, que permite à empresa visualizar onde os clientes estão abandonando o processo de compra. O uso de Lambda para essa tarefa permite atualizações diárias dos dashboards, garantindo que as decisões sejam baseadas em dados atualizados.

As Lambda Functions tem um papel central na arquitetura de *big data*, A combinação do modelo *serverless* com a capacidade de particionamento e transformação dos dados demonstra como as Lambda Functions são uma escolha interessante para construir uma arquitetura de *big data* robusta e escalável, capaz de suportar as operações de empresas de telecomunicações e de oferecer *insights* rápidos e precisos para a otimização contínua de suas operações de vendas automáticas.

Para orquestrar todas essas Lambda Functions de forma eficiente e garantir que cada uma seja executada na ordem correta, foi utilizado o serviço AWS Step Functions e podemos ver sua definição na Figura 3. Step Functions é um serviço de orquestração de fluxo de trabalho que facilita a coordenação de múltiplos serviços AWS em fluxos de trabalho visuais, oferecendo um controle detalhado sobre a sequência de execução e a lógica condicional entre as etapas.

Ao integrar o AWS Step Functions na arquitetura, a equipe conseguiu orquestrar as Lambda Functions de maneira eficiente e confiável, garantindo que cada etapa do processo de dados fosse executada corretamente e na sequência adequada. Essa integração simplificou a gestão da orquestração e também aumentou a robustez e a escalabilidade da

solução de *big data*, permitindo que a empresa operasse de forma mais ágil e responsiva às necessidades de negócio, adicionando e removendo componentes do fluxo de trabalho.

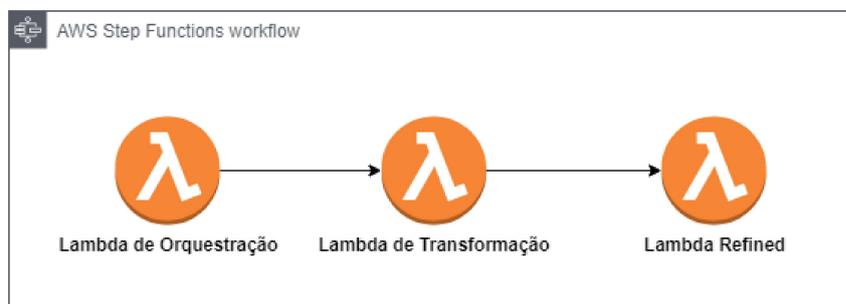


Figura 3 – Orquestração AWS StepFunction

Fonte - Própria

Para facilitar a visualização, o funil foi dividido em macro etapas da jornada de compra. Cada Macro etapa era dividida de duas a três micro etapas que cada uma era associada ao número de usuários que atingiu essa etapa conforme apresentado na Figura 4.

O funil de vendas gerado pela Lambda *Refined* tornou-se uma ferramenta essencial para a equipe de análise de dados. Ele permitiu que a empresa visualizasse, as etapas específicas em que os clientes estavam abandonando o processo de compra. Esse nível de detalhame foi crucial para identificar problemas específicos, como bugs em determinadas etapas ou desinteresse nas ofertas apresentadas.



Figura 4 – Funil de vendas separado por macro etapas

Fonte - Própria

Com essas informações, a equipe foi capaz de direcionar seus esforços para solucionar problemas específicos no fluxo de vendas, seja ajustando as ofertas, corrigindo falhas técnicas ou melhorando a experiência do usuário. Como resultado, houve a melhora da eficiência do *chatbot* e também o aumento a satisfação dos clientes, gerando uma experiência de compra mais fluida e agradável.

## 5 Resultados

Este estudo de caso exemplifica a transformação de uma empresa no ramo de telecomunicações que adotando uma cultura orientada a dados, conseguiu superar desafios significativos, principalmente na otimização de suas vendas automatizadas. O desenvolvimento de uma arquitetura de *big data* em nuvem possibilitou uma visibilidade detalhada do comportamento dos usuários ao longo da jornada de compra. Esse processo de mudança foi fundamental para a compreensão, fundamentada em dados, das razões pelas quais um usuário pode não se transformar em um consumidor ao fim de uma jornada de compra.

Em comparação ao desempenho do *chatbot* pré implementações, quando não havia uma estrutura organizada para coletar e analisar dados, os ganhos foram notáveis. No período anterior, as decisões eram frequentemente baseadas em suposições ou percepções subjetivas, levando a uma ineficiência generalizada. Os problemas no fluxo de vendas eram abordados de maneira dividida, com cada setor buscando culpados sem dados concretos para fundamentar suas decisões. Esse cenário mudou drasticamente em um cenário pós mudança com a implementação de um *dashboard* que consolidou as informações de cada etapa da jornada de compra. Essa mudança permitiu uma tomada de decisão mais precisa e, conseqüentemente, melhorias operacionais e financeiras.

A arquitetura de *big data*, desenvolvida com foco na escalabilidade e eficiência, permitiu que a equipe monitorasse o desempenho do *chatbot* de vendas. Um dos exemplos mais claros dessa melhoria foi a identificação de um *bug* que afetava a conversão na etapa de pagamento por "Débito Automático". Antes, esse tipo de problema passava despercebido por longos períodos, o que impactava as vendas. Com a nova estrutura de dados, foi possível identificar e corrigir o *bug* de forma ágil, minimizando as perdas financeiras associadas a essa falha. Esse é apenas um exemplo da agilidade que o uso de dados guiou a equipe.

Outro ponto positivo foi a melhoria na conversão entre as etapas de "Selecionar Produto" e "Selecionar CEP". A baixa taxa de conversão foi rapidamente detectada, e com base nos dados, a equipe optou por reformular o esse menu. Essa alteração, validada pela análise posterior, resultou em um aumento considerável na conversão, evidenciando como o uso de dados permitiu ajustes precisos e eficazes. Esse ciclo de validação e ajuste constante aprimorou o desempenho do *chatbot* como e gerou confiança nas decisões tomadas pela equipe.

Entretanto, ao longo do processo de adaptação à nova infraestrutura, alguns desafios surgiram. Um dos problemas observados foi a rigidez inicial na construção da tabela

*refined*, utilizada para alimentar o dashboard. A cada alteração no fluxo do funil de vendas, seja a adição, remoção ou mudança na ordem das etapas, era necessário reconstruir a tabela, o que aumentava significativamente o tempo de processamento e dificultava a agilidade nas tomadas de decisão. Essa limitação gerava atrasos na implementação de ajustes impactava na eficiência do produto além de impactos financeiros.

Com o tempo, a equipe identificou que não era necessário manter a ordem das etapas no nível da tabela *refined*. Em vez disso, a ordem poderia ser ajustada na visualização final do dashboard, eliminando a necessidade de reconstruir a tabela a cada mudança no funil de vendas. Essa simples mudança de abordagem foi importante para aumentar a flexibilidade do sistema e garantir que as alterações pudessem ser implementadas de forma mais rápida e eficiente. Esse ajuste, embora tenha surgido de um desafio, exemplifica a capacidade da solução ser flexível e evoluir durante o processo.

Outro aspecto importante a ser mencionado é o impacto financeiro dessa transformação, ainda que os números não possam ser divulgados explicitamente. A partir do momento em que a empresa passou a basear suas decisões em dados precisos, foi possível otimizar recursos e direcionar investimentos de forma mais estratégica. O aumento da eficiência nas vendas automatizadas, em conjunto com a redução de falhas e a evolução contínua do funil de vendas, trouxe uma melhora significativa nos indicadores de desempenho. Essas melhorias refletiram diretamente nos resultados financeiros da empresa, que experimentou um crescimento importante em relação ao períodos anteriores.

Apesar dos desafios, o processo de transição para uma cultura orientada a dados se mostrou altamente benéfico. Observou-se a melhora da eficiência operacional e também conseguiu a adaptação das suas operações de forma mais ágil e precisa. Essa capacidade de adaptação foi essencial para que a equipe de desenvolvimento pudesse evoluir o produto de vendas automáticas e solucionar os problemas à medida que surgiam.

Para o futuro, as lições aprendidas com a implementação da arquitetura de *big data* e a transformação da cultura organizacional são valiosas. Uma infraestrutura tecnológica robusta, capaz de lidar com grandes volumes de dados entregou a possibilidade de se conhecer melhor os usuários e suas razões para não se tornarem consumidores. Além disso, a flexibilidade conquistada no processo de gestão dos dados permite que a equipe continue a otimizar o *chatbot* sempre com base em *insights* claros, precisos e fundamentados em dados.

## 6 Conclusão

Em conclusão, este estudo de caso exemplifica como a adoção de uma cultura orientada a dados, apoiada por uma infraestrutura tecnológica adequada, pode transformar profundamente uma empresa. Os ganhos financeiros, operacionais e estratégicos, embora não possam ser revelados em detalhes, são evidentes nas melhorias de desempenho e eficiência obtidas.

Com trabalho de conclusão de curso evidenciou-se a importância da implementação de uma arquitetura de *big data* para a otimização de processos de vendas automatizadas. Considerando desafios técnicos, operacionais e culturais, todos tratados com uma abordagem orientada a dados e uma infraestrutura tecnológica robusta. Ao fim, tem-se uma solução técnica exemplar, eficaz e aprendizados interessantes que podem ajudar futuros profissionais que buscam transformar seu contexto com o uso de dados.

Um dos desafios enfrentados foi a falta de flexibilidade inicial no processo de construção das tabelas *refined*. Para os profissionais que encontrarem algo semelhante, um caminho é evitar estruturas rígidas que dependam de reconstruções toda vez que uma mudança no fluxo for necessária. Produtos digitais são extremamente dinâmicos e é importante olhar para esses dinamismos de forma positiva pois é ele que abre margem para melhorias. A flexibilidade na estrutura dos dados é fundamental para garantir agilidade nas respostas às mudanças. No caso apresentado aqui, a alteração na forma como as etapas do funil de vendas eram organizadas no *dashboard* trouxe um ganho em termos de eficiência. A estrutura de dados não deve ser um obstáculo, mas uma facilitadora para ajustes rápidos e contínuos.

Outro aspecto importante a destacar é a importância de olhar para os erros como parte natural do processo de inovação. O que foi apresentado aqui não é apenas uma exemplificação de um caso de sucessos, mas também de falhas e ajustes. Por exemplo, durante a implementação, percebeu-se que a adição ou remoção de etapas no funil de vendas impactava negativamente a observabilidade do produto, atrasando a capacidade da equipe de tomar decisões rápidas. Esse erro levou a repensar a estrutura da tabela *refined*, permitindo que ela fosse ajustada no dashboard sem precisar ser reconstruída a cada mudança. Essa lição de flexibilidade e aprendizado contínuo será fundamental para quem assumir o próximo desafio.

Uma das contribuições do presente estudo é exemplificar o poder da análise de dados quando bem aplicada, pois ficou evidente que decisões informadas por dados são mais precisas, ágeis e eficazes. O sucesso do produto no aprimoramento de seu *chatbot* de vendas não veio apenas de tecnologia, mas da capacidade de transformar dados brutos

em informações sobre o usuário e informações em ações justificadas. Esse processo é algo que outros profissionais podem replicar em outros setores de atividades e diferentes tipos de negócios.

Para quem continuar este caminho, é preciso valorizar a importância de uma infraestrutura de dados bem planejada e todo o aspecto tecnológico, mas também de estar preparado para ajustes contínuos. A evolução de produtos digitais, orientada por dados, não é um destino final, mas sim um passo na melhoria contínua.

# Referências

ASSIS, R. M. d. et al. Preparando bases de dados para uso em sistema preditivo que visa a reduzir a emissão de ordens de serviço em empresa de telecom. Universidade Federal de Uberlândia, 2022. Citado na página 9.

BEZERRA, C. A. V.; BEZERRA, C. D. d. S.; PAGOTTI, L. F. et al. Digitalização na indústria 4.0-análise de digitalização de fábricas: Implementação de sistemas de monitoramento, coleta de dados e automação. Pontifícia Universidade Católica de Goiás, 2023. Citado na página 8.

BROWN, T. Change by design: How design thinking transforms organizations and inspires innovation.[kindle 2 version]. **Retrieved from Amazon. com**, 2009. Citado na página 14.

BUSTAMANTE, R. S. A. Adoção de solução de cloud computing na área de negócios de uma fábrica de fármacos e cosméticos. Universidade Federal de Uberlândia, 2022. Citado na página 9.

DICK, A. S.; BASU, K. Customer loyalty: toward an integrated conceptual framework. **Journal of the academy of marketing science**, Springer, v. 22, p. 99–113, 1994. Citado na página 8.

FADER, P. **Customer centricity: Focus on the right customers for strategic advantage**. [S.l.]: University of Pennsylvania Press, 2020. Citado na página 15.

FOSTER, I.; GANNON, D. B. **Cloud computing for science and engineering**. [S.l.]: MIT Press, 2017. Citado na página 16.

GARRETT, J. J. Customer loyalty and the elements of user experience. **Design management review**, v. 17, n. 1, 2006. Citado na página 14.

KARTAJAYA, H.; KOTLER, P.; HOOL, D. H. Marketing 4.0: moving from traditional to digital. **World Scientific Book Chapters**, World Scientific Publishing Co. Pte. Ltd., p. 99–123, 2019. Citado na página 14.

KELLER, K. L. Conceptualizing, measuring, and managing customer-based brand equity. **Journal of marketing**, Sage Publications Sage CA: Los Angeles, CA, v. 57, n. 1, p. 1–22, 1993. Citado na página 8.

KIMBALL, R.; CASERTA, J. **The data warehouse ETL toolkit**. [S.l.]: John Wiley & Sons, 2004. Citado 2 vezes nas páginas 17 e 18.

MADEIRA, A. C. M.; NEVES, B. C.; JESUS, B. Daniel de. O uso da inteligência artificial aplicada ao marketing digital: exploração das vulnerabilidades do usuário-consumidor (the use of artificial intelligence applied to digital marketing: exploitation of user-consumer vulnerabilities). **Journal of Digital Media & Interaction**, Vol. 3, No. 8, University of Aveiro, v. 3, n. 8, p. 95–111, 2020. Citado na página 12.

- MELL, P.; GRANCE, T. et al. The nist definition of cloud computing. Computer Security Division, Information Technology Laboratory, National ... , 2011. Citado na página 15.
- MENDES, T. d. R. O. **Comportamento do consumidor**. Tese (Doutorado), 2014. Citado na página 8.
- NAMBIAR, A.; MUNDRA, D. An overview of data warehouse and data lake in modern enterprise data management. **Big data and cognitive computing**, MDPI, v. 6, n. 4, p. 132, 2022. Citado 2 vezes nas páginas 16 e 17.
- NORMAN, D. **The design of everyday things: Revised and expanded edition**. [S.l.]: Basic books, 2013. Citado 4 vezes nas páginas 11, 12, 13 e 14.
- OLIVEIRA, M. H. B. d. Práticas para garantia de qualidade em produtos digitais. Universidade Federal de Uberlândia, 2023. Citado na página 9.
- OLIVER, R. L. **Satisfaction: A behavioral perspective on the consumer: A behavioral perspective on the consumer**. [S.l.]: Routledge, 2014. Citado na página 8.
- RITTINGHOUSE, J. W.; RANSOME, J. F. **Cloud computing: implementation, management, and security**. [S.l.]: CRC press, 2017. Citado 3 vezes nas páginas 15, 16 e 17.
- ROZADOS, H. F.; PIFFER, B. P. Pesquisa de marketing e estudos de usuário: um paralelo entre os dois processos. **Em Questão**, Universidade Federal do Rio Grande do Sul, v. 15, n. 2, p. 169–182, 2009. Citado na página 8.
- SOLOMON, M. R. **O Comportamento do consumidor-: comprando, possuindo e sendo**. [S.l.]: Bookman Editora, 2016. Citado 2 vezes nas páginas 11 e 13.