



**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DA PARAÍBA
CAMPUS MONTEIRO**

**DIRETORIA DE DESENVOLVIMENTO DE ENSINO
TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS**

GUILHERME RODRIGUES DA COSTA

**RELATÓRIO DE ESTÁGIO
ESTÁGIO EM ENGENHARIA DE DADOS NA EMPRESA COMPASS UOL**

MONTEIRO

2024

GUILHERME RODRIGUES DA COSTA

RELATÓRIO DE ESTÁGIO
ESTÁGIO EM ENGENHARIA DE DADOS NA EMPRESA COMPASS UOL

Relatório de Estágio apresentado à Coordenação de Estágio do Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, Campus Monteiro, como requisito parcial para conclusão do Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas

Orientador: Prof. Me. Giuseppe Anthony Nascimento de Lima.

MONTEIRO
2024

Dados Internacionais de Catalogação na Publicação – CIP
Bibliotecária responsável Porcina Formiga dos Santos Salgado CRB15/204
IFPB Campus Monteiro.

C837e Costa, Guilherme Rodrigues da.

Estágio em Engenharia de Dados na Empresa COMPASS-UOL /
Guilherme Rodrigues da Costa – Monteiro-PB. 2023.
34fls. : il.

Relatório (Curso Superior de Tecnologia em Análise e
Desenvolvimento de Sistemas) - Instituto Federal de Educação,
Ciência e Tecnologia da Paraíba - IFPB campus, Monteiro.

Orientador: Prof. Me. Giuseppe Anthony Nascimento de Lima.

1. Dados - engenharia 2. Big data 3. Empresa COMPASS UOL
I. Título .

CDU 004.422.6

GUILHERME RODRIGUES DA COSTA

**RELATÓRIO DE ESTÁGIO
OBTENDO EXPERIÊNCIA EM ENGENHARIA DE DADOS NA EMPRESA
COMPASS UOL**

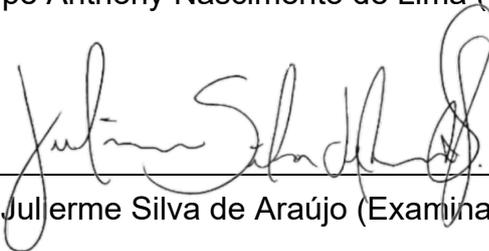
Relatório de Estágio apresentado à
Coordenação de Estágio do Instituto
Federal de Educação, Ciência e
Tecnologia da Paraíba, Campus Monteiro,
como requisito parcial para conclusão do
Curso Superior de Tecnologia em Análise
e Desenvolvimento de Sistemas

Aprovado em 01 de abril de 2024.

BANCA EXAMINADORA



Prof. Me. Giuseppe Anthony Nascimento de Lima (Orientador – IFPB)



Prof. Me. Julierme Silva de Araújo (Examinador – IFPB)



Prof. Esp. Wagner de Oliveira Santos (Examinador – IFPB)

AGRADECIMENTOS

Quero expressar minha profunda gratidão a todas as pessoas que contribuíram para a realização deste relatório de estágio, tornando esta jornada uma experiência significativa e enriquecedora.

Em especial, quero agradecer à minha mãe, cujo apoio constante e encorajamento foram fundamentais ao longo de toda a trajetória. Sua dedicação e incentivo foram minha fonte de força nos momentos desafiadores, e sua presença amorosa é o pilar sobre o qual construí meus sucessos.

Além disso, estendo meus agradecimentos a todos os colegas do estágio, supervisores, professores e amigos que compartilharam seus conhecimentos, experiências e conselhos valiosos durante o período de estágio. Cada interação e aprendizado moldaram não apenas este relatório, mas também minha visão profissional e pessoal.

Agradeço a todos que, de alguma forma, contribuíram para o meu crescimento e aprendizado. Este relatório é resultado não apenas do meu esforço, mas do apoio generoso de uma rede de pessoas incríveis. Muito obrigado a cada um de vocês.

RESUMO

A engenharia de dados é uma área fundamental para empresas de todos os setores, uma vez que seus profissionais são responsáveis por coletar, armazenar, processar e analisar dados, transformando-os em informações valiosas que podem ser utilizadas para tomar decisões estratégicas. Portanto, a demanda por engenheiros de dados tem aumentado, em que faltam profissionais qualificados. Este trabalho relata a experiência de estágio na empresa Compass UOL, desde a aprovação em um processo seletivo de vagas com bolsa, que envolveu o contato com ferramentas de controle de versão de software e de extração, transformação e carregamento de dados (ETL). O estágio também proporcionou uma maior proficiência nas linguagens programação Python e SQL e uma compreensão de conceitos da área de Big Data, abrangendo a prática em gerenciamento de serviços em nuvem para análise de dados, com ênfase no ecossistema Amazon Web Services (AWS).

Palavras-chave: estágio; engenharia de dados; *big data*; análise e desenvolvimento de sistemas.

ABSTRACT

Data engineering is a fundamental area for companies in all sectors, as it's professionals are responsible for collecting, storing, processing and analyzing data, transforming them into valuable information that can be used to make strategic decisions. Therefore, the demand for data engineers has increased in which there is a lack of qualified professionals. This work reports the internship experience at the company Compass UOL, since the selection process for scholarship positions, which involved contact with software version control and data extraction, transformation and loading (ETL) tools. The internship also increased the proficiency in Python and SQL programming languages and the understanding of concepts in Big Data area, covering the practice of managing cloud services for data analysis, with an emphasis on the Amazon Web Services (AWS) ecosystem.

Keywords: internship; data engineering; big data, system analysis and development.

LISTA DE FIGURAS

Figura 1 – Logomarca da Compass UOL.....	13
Figura 2 – Pilares de Tecnologia da Compass UOL.....	14
Figura 3 – Distribuição de Localidades da Compass UOL.....	14
Figura 4 – Representação do Processo Scrum para Planejamento de Sprint.....	17
Figura 5 – Relatórios com a Pontuação do Feedback Quinzenal.....	18
Figura 6 – Organização do Repositório Git.....	19
Figura 7 – Dashboard Carregado com Dados Pós-Processados em ETL no Excel.....	22
Figura 8 – Diferença do Docker e a Virtualização Tradicional.....	26
Figura 9 – Serviços da Amazon Web Services (AWS).....	27
Figura 10 – Console da Amazon Web Services (AWS).....	28
Figura 11 – Dashboard TMDB com AWS Quicksight.....	30

LISTA DE ABREVIATURAS

API	<i>Application Programming Interface</i> (Interface de Programação de Aplicação)
AWS	<i>Amazon Web Services</i>
CST	Curso Superior de Tecnologia
CSV	<i>Comma-Separated-Values</i> (Valores Separados por Vírgulas)
D2E	<i>Data-Driven Everything</i>
EC2	<i>Amazon Elastic Compute Cloud</i>
ETL	<i>Extract, Transform and Load</i> (Extração, Transformação e Carregamento)
IAAS	<i>Infrastructure-as-a-Service</i> (Infraestrutura como Serviço)
IBM	<i>International Business Machines</i>
IOT	<i>Internet of Things</i> (Internet das Coisas)
MVP	<i>Minimum Viable Product</i> (Mínimo Produto Viável)
OO	Orientado a Objeto
PAAS	<i>Platform as a Service</i> (Plataforma como Serviço)
RDD	<i>Resilient Distributed Datasets</i>
REST	<i>Representational State Transfer</i> (Transferência de Estado Representacional)
SAAS	<i>Software as a Service</i> (Software como Serviço)
SO	Sistema Operacional
SQL	<i>Structured Query Language</i>
S3	<i>Simple Storage Service</i>
TMDB	<i>The Movie Database</i>
VM	<i>Virtual Machine</i>

SUMÁRIO

1 INTRODUÇÃO.....	9
2 OBJETIVOS.....	10
2.1 Objetivo geral.....	10
2.2 Objetivos específicos.....	10
3 RELATO DE ESTÁGIO.....	13
3.1 Identificação e perfil da empresa.....	13
3.1.1 Local do estágio e aspectos físicos da empresa.....	14
3.1.2 Dados referentes ao estágio.....	15
3.2 Atividades desenvolvidas.....	15
3.2.1 Seleção, preparação e acompanhamento do estágio remoto.....	15
3.2.2 Gerenciamento de versões com Git e GitHub.....	19
3.2.3 Aprofundamento no Sistema Operacional Linux.....	20
3.2.4 Processo de manipulação de dados.....	21
3.2.6 Aperfeiçoamento na linguagem Python.....	24
3.2.7 Implementação de Ambiente Docker e Uso do Apache Spark.....	25
3.2.8 Utilização dos serviços AWS.....	26
3.2.9 Análise de Dados, Modelagem para Normalização, Dimensionamento e Multidimensionalização dos Dados.....	28
4 CONSIDERAÇÕES FINAIS.....	31
REFERÊNCIAS.....	32

1 INTRODUÇÃO

No cenário tecnológico atual, o desenvolvimento de software e a engenharia de dados são fundamentais para impulsionar a inovação e a eficiência nas organizações, considerando as tecnologias da informação e comunicação.

De acordo com o artigo publicado pela jornalista Júlia Nunes (2023) no portal de notícias G1, as empresas com bancos, seguradoras, saúde e entre outras, estão em uma alta demanda a procura de profissionais de tecnologia. O mesmo destaca que um engenheiro de dados é aquele que faz a coleta dos dados e organiza as informações dentro da estrutura que foi determinada pelo arquiteto de dados. Como também prevê uma tendência salarial para a área de engenharia de dados em 2024, estimando uma média de R\$ 24,1 mil. Essa informação destaca a valorização crescente do engenheiro de dados diante da relevância estratégica que a manipulação de dados assume nos setores mencionados.

Conforme a IBM (2023) destaca, os desenvolvedores de software não apenas desempenham um papel vital na codificação, mas também são protagonistas no ciclo de vida completo de um software. Sua atuação abrange desde a tradução de requisitos em recursos tangíveis até a coordenação entre equipes funcionais e a garantia da qualidade do produto de software final.

A gestão eficaz de projetos de software recorrentemente é um desafio crítico nesta área. Há uma preocupação desafiadora em entregá-los dentro do prazo e orçamento preestabelecidos, de garantir a sintonia com os objetivos organizacionais e de cultivar uma cultura ágil são aspectos (MARTINS, 2023). Tais desafios reforçam a importância de uma abordagem dinâmica e inovadora na gestão de projetos de software.

A engenharia de dados é um domínio crucial para a transformação de dados brutos em informações úteis e acionáveis (BELMONTE, 2023). Os engenheiros de dados desempenham um papel vital na concepção, construção e manutenção de infraestruturas de dados escaláveis e seguras, garantindo assim a qualidade e disponibilidade dos dados (UNIFOR, 2023). Esses aspectos são essenciais para análises precisas e fundamentadas, que são a base das tomadas de decisão nas mais diversas organizações.

A partir de divulgação de oportunidades de estágio do CST em Análise e Desenvolvimento de Sistemas, com a intenção de capacitação e primeira imersão e

engenharia de dados, participei do processo seletivo de vagas de estágio com bolsa da empresa de tecnologia Compass UOL, que abrangeu avaliações de proficiência em português e inglês, além de um desafio lógico que avaliou habilidades analíticas.

Neste relatório de estágio são detalhadas as atividades desempenhadas nesse processo, destacando-se a participação em times de projetos de implementação de infraestruturas de dados, o desenvolvimento e aprimoramento de códigos, bem como a interação com diversas ferramentas e métodos específicos da área.

2 OBJETIVOS

2.1 Objetivo geral

Este relatório de estágio objetivou registrar o processo de aquisição de conhecimentos práticos em ferramentas e técnicas de engenharia de dados, partindo-se daqueles já obtidos na graduação no CST em Análise e Desenvolvimento de Sistemas, considerando a imersão efetiva em um ambiente colaborativo de capacitação e produção de software na empresa Compass UOL.

2.2 Objetivos específicos

- Compreender o uso de ferramentas de controle de versão: aquisição de habilidades avançadas no uso do Git e GitHub para controle de versão e colaboração eficaz em projetos de software.
- Adotar o ambiente Linux: aplicação de comandos sobre o sistema de arquivos, de instalação e manutenção de pacotes e de controle de processos, de forma a aplica-lo como sistema operacional em ambientes de execução de Engenharia de Dados.
- Aprimorar a manipulação de dados com a linguagem SQL: competência avançada em escrever consultas SQL complexas para extrair, transformar e carregar dados de diferentes fontes.
- Compreender conceitos de Big Data: adquirir conhecimento abrangente sobre aplicabilidade de tecnologias envolvidas no processamento de grandes volumes de dados.
- Programar com a linguagem Python: uso mais avançado de Python, incluindo o contato com bibliotecas como Pandas, Numpy e Apache Spark para análise e manipulação de dados.
- Gerenciar o ambiente Docker para a implantação de aplicações de software: competência em criar e gerenciar contêineres Docker para facilitar a implantação de ambientes de execução de aplicações em Engenharia de Dados, em sistemas Linux.

- Explorar e usar serviços AWS (Athena, Glue, S3 e Lambda): contato com serviços de computação em nuvem providos pela AWS, incluindo Athena para consultas em dados no S3; Glue para ETL (*Extract, Transform, Load*) automatizado de dados; Lambda para execução de código sem provisionamento de servidor; e S3 para armazenamento eficiente e seguro.
- Projeto de análise e visualização de dados: implementação de solução a partir do consumo de uma API REST provida por um repositório aberto de dados sobre filmes, com tratamento para exibi-los em uma *dashboard* com *insights* analíticos.

3 RELATO DE ESTÁGIO

3.1 Identificação e perfil da empresa

A Compass UOL é uma empresa que atua no segmento de tecnologia da informação (TI), oferecendo serviços inovadores e produtos digitais. Com uma estrutura organizacional flexível e dinâmica, a empresa destaca-se por sua capacidade de adaptação às mudanças do mercado e pela entrega de soluções tecnológicas avançadas.

Fundada em 1995, a empresa anteriormente denominada CompassoUOL, passou por uma transição de nome durante seu processo de expansão global, marcando sua entrada nos mercados norte-americano e europeu. Essa mudança para a marca Compass UOL (Figura 1) reflete a estratégia de crescimento internacional da empresa, conforme ressaltado em seu blog (COMPASS UOL, 2021), destacando-se no compromisso com a transformação digital e na implementação de soluções avançadas, incluindo o uso de inteligência artificial (IA) empresarial.

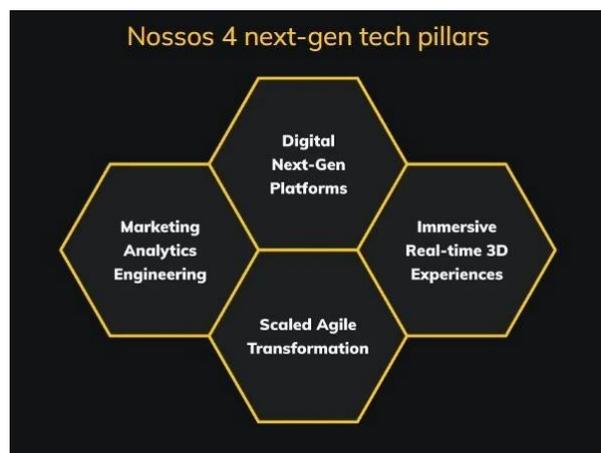
Figura 1 – Logomarca da Compass UOL



Fonte: (COMPASS UOL, 2024a).

A empresa dedica seus esforços ao desenvolvimento e construção de plataformas nativas digitais, com foco na inovação tecnológica e na resposta às necessidades de empresas líderes em escala global. Para impulsionar esse crescimento digital, são estabelecidos quatro pilares de conhecimento distribuídos entre 14 *Innovation Studios*, com o objetivo de criar experiências inovadoras e modelos de negócios disruptivos em todos os setores (Figura 2).

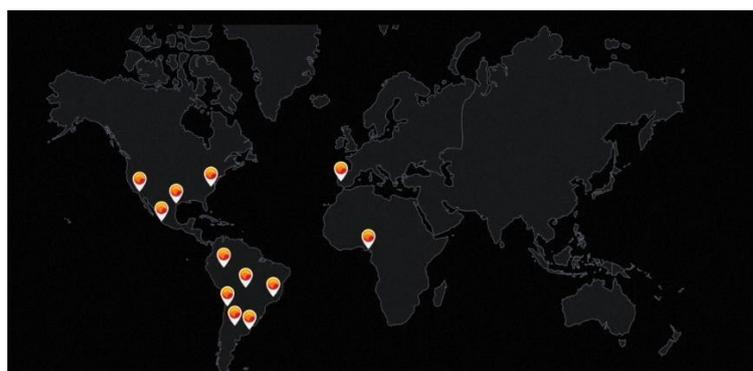
Figura 2 – Pilares de Tecnologia da Compass UOL



Fonte: (COMPASS UOL, 2024b).

Para manter uma presença global significativa, a Compass UOL adota uma estratégia abrangente, estabelecendo sedes em diversas regiões do mundo. Essa expansão geográfica permite à empresa adaptar-se às especificidades locais, fortalecendo sua influência nos mercados internacionais. Além disso, essa abordagem estratégica será detalhada na Figura 3, ilustrando de maneira visual a distribuição geográfica das sedes e evidenciando o compromisso da Compass UOL em ser uma força global em seu setor.

Figura 3 – Distribuição de Localidades da Compass UOL



Fonte: (COMPASS UOL, 2024c).

Entre os produtos e serviços da empresa, destaca-se a parceria estratégica com a Amazon Web Services (AWS), conforme evidenciado no case do Sicredi D2E (COMPASS UOL, 2024d). Essa parceria resultou no desenvolvimento de um *roadmap*

para recomendação do produto mínimo viável (MVP) e na implementação de melhorias contínuas.

Em conjunto com a AWS, a empresa concentra esforços na construção de modelos de segmentação comportamental dos associados e de propensão à evasão, utilizando dados históricos de investimentos. Essa iniciativa, notabilizada pelo “Workshop de Ideação e Design” do programa AWS Data-Driven Everything (AWS D2E), tem como objetivo extrair *insights* dos dados para recomendar produtos de investimentos e oferecer atendimento preventivo para retenção, contribuindo assim para a estratégia de crescimento digital da Compass UOL e fortalecendo a posição da empresa no cenário de transformação digital.

3.1.1 Local do estágio e aspectos físicos da empresa

A Compass UOL possui sede fisicamente localizada em Passo Fundo, Rio Grande do Sul. Os estagiários ingressantes de seu programa de estágio são distribuídos em *squads* remotas (equipes), que assumem a responsabilidade pelo desenvolvimento de projetos ligados às operações da empresa, interagindo por meio de videoconferências e ferramentas de desenvolvimento e comunicação disponibilizadas pela empresa.

Próximo à eclosão da pandemia de COVID-19, a Compass UOL já se encontrava em um processo de transição do trabalho presencial para o remoto. Inicialmente, a empresa adotou um modelo híbrido, combinando dias de trabalho presencial com dias remotos. Contudo, a pandemia acelerou consideravelmente essa transição (assim como em outras empresas), resultando em um cenário em que todo o trabalho pode ser realizado de forma remota ou presencial, conforme a preferência individual e o tipo de colaboração nos projetos.

Nesse sentido, a Compass UOL dispõe de diversas ferramentas de TI para facilitar a comunicação e a colaboração de suas equipes. No estágio, a empresa disponibilizou ferramentas colaborativas e de comunicação da suíte de escritório da Microsoft (Microsoft Teams¹, Outlook², entre outros). Na comunicação assíncrona o e-

¹ Disponível em: <https://www.microsoft.com/pt-br/microsoft-teams/group-chat-software>. Acesso em: 30 dez. 2023.

² Disponível em: <https://www.microsoft.com/pt-br/microsoft-365/outlook/email-and-calendar-software-microsoft-outlook>. Acesso em: 15 jan. 2024.

mail é utilizado, mas a comunicação cotidiana é viabilizada por meio de videoconferência (internamente, chamadas de “call”), considerando uma gestão de atividades e projetos de forma remota.

Durante o período do estágio, a empresa disponibilizou o acesso a serviços em nuvem de terceiros, a partir da infraestrutura da Amazon Web Service (AWS), para a realização das atividades.

3.1.2 Dados referentes ao estágio

O estágio teve início no dia 02 de agosto de 2023 e término em 29 de dezembro de 2023, totalizando 400h. O regime de carga-horária foi 20 horas semanais, de segunda à sexta, das 13:30 às 17:30.

3.2 Atividades desenvolvidas

3.2.1 Seleção, preparação e acompanhamento do estágio remoto

O processo de seleção para o estágio remoto foi conduzido por meio da plataforma de recrutamento GUPY³ e compreendeu três avaliações escritas abrangentes, em português, em inglês e de raciocínio lógico. Após a aprovação, foram coordenados os procedimentos para a apresentação de documentos e a realização do exame admissional na cidade de Monteiro - PB.

Foram disponibilizadas contas de acesso ao e-mail corporativo e também às plataformas da Microsoft e da Udemy Business⁴. Essa última proporcionou o acesso aos cursos listados a seguir, todos relacionados com a formação exigida na experiência de estágio, em que foram obtidos certificados de conclusão deles:

- Git e GitHub do básico ao avançado (c/ gist e GitHub Pages)
- Linux para Desenvolvedores (c/ terminal, Shell, Apache e +)
- SQL para Análise de Dados - do Básico ao Avançado
- Python 3 - Curso Completo do Básico ao Avançado
- Docker para Desenvolvedores (com Docker Swarm e Kubernetes)

³ Disponível em: <https://www.gupy.io>. Acesso em: 15 jan. 2024.

⁴ Disponível em: <https://www.udemy.com/>. Acesso em: 15 jan. 2024.

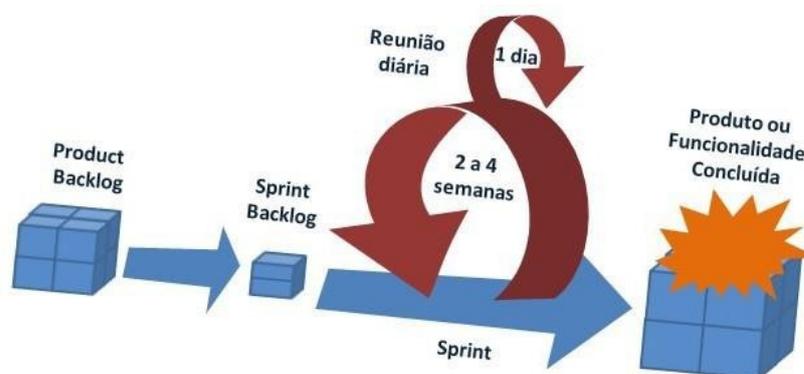
- Estatística Descritiva com Python
- Learn By Example: Hadoop, MapReduce for Big Data problems
- Formação Spark com Pyspark: o Curso Completo

Ainda, foi disponibilizada uma conta de acesso para a Amazon Web Services (AWS), ao longo do estágio.

No primeiro contato foi explicado o objetivo da empresa de capacitar pessoal na área de engenharia de dados, visando posterior contratação. As atividades do estágio eram designadas em iterações quinzenais (*sprints*), em que deveriam ser realizadas entregas de software que exigiram estudos em tecnologias e contato com ferramentas de desenvolvimento.

Metodologicamente, cada *sprint* começava com a apresentação das entregas que eram esperadas e de que fontes de estudo eram necessárias para produzi-las, equivalentemente ao *sprint planning* da metodologia ágil Scrum (SCHWABER; SHUTERLAND, 2020). Essa abordagem estruturada em *sprints* proporcionou uma visão clara dos objetivos e direcionou as atividades a serem realizadas durante cada iteração (Figura 4).

Figura 4 – Representação do Processo Scrum para Planejamento de Sprint



Fonte: (MIND MASTER TREINAMENTOS, 2024).

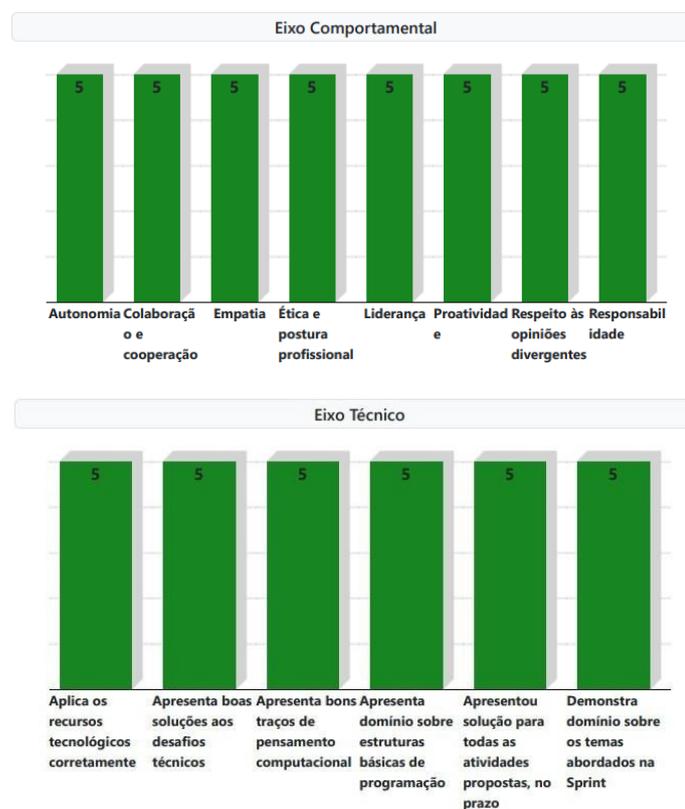
Diariamente, ocorria o equivalente ao *daily scrum*, em que um líder experiente da empresa acompanhava cada estagiário de uma mesma equipe (os estagiários foram distribuídos em times ou *squads*), verificando o que conseguiram fazer no dia anterior e o que pretendiam fazer no dia corrente, além de serem compartilhadas as

dificuldades (impedimentos técnicos) e respectivas estratégias de solução. Fora dessas reuniões diárias, houve um acompanhamento de forma assíncrona (mensagens ou videoconferência), por meio do contato com o líder, em caso de dúvidas.

Cada *sprint* era concluída com o evento *sprint review* do Scrum, em que um quadro Kanban era construído para cada membro da equipe, considerando aspectos positivos, negativos e tomadas de ações de melhorias para a próxima iteração, considerando o desempenho do membro na equipe.

Esse processo subsidiou o acompanhamento das habilidades pessoais, interpessoais (*soft skills*) e técnicas (*hard skills*) de cada estagiário pela empresa, em que foi utilizado um sistema de *feedback* quinzenal (término da *sprint*, equivalendo a uma *sprint review* no Scrum). No caso, essa avaliação era enviada por e-mail aos estagiários, englobando a avaliação de aspectos como comunicação, empatia, proatividade, prazo e qualidade da entrega, entre outros (Figura 5).

Figura 5 – Relatórios com a Pontuação do Feedback Quinzenal



Fonte: próprio autor.

3.2.2 Gerenciamento de versões com Git e GitHub

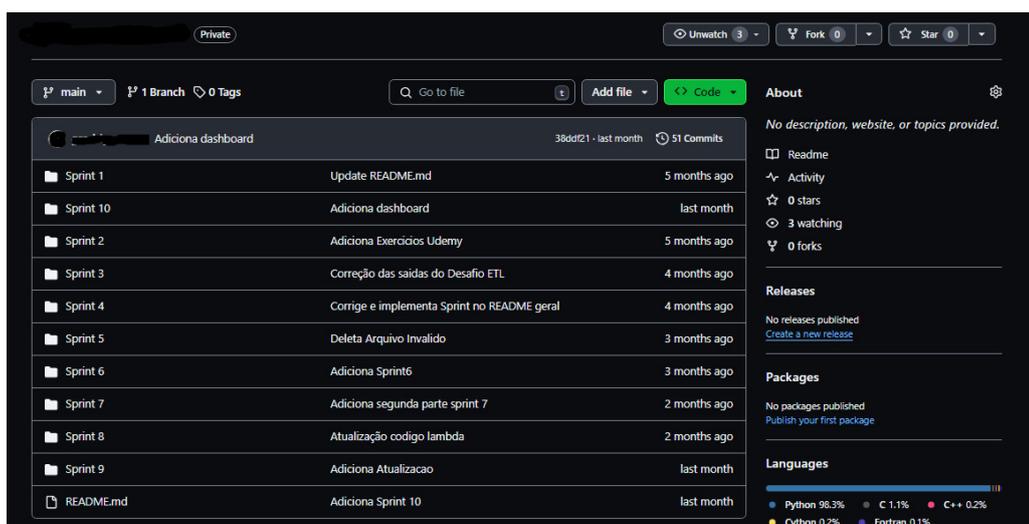
A gestão eficiente das alterações em código fonte é fundamental no desenvolvimento de software e os sistemas de controle de versão desempenham um papel central nesse processo. Essas ferramentas automatizadas registram e gerenciam as modificações em um código ao longo do tempo, criando instantâneos dos arquivos durante o desenvolvimento (MICROSOFT, 2023).

Essa abordagem evita problemas associados com a falta de controle de versão, permitindo que a equipe trabalhe colaborativamente de forma mais segura, mantendo-se históricos precisos de mudanças.

O estágio disponibilizou a realização de um curso abrangente sobre as ferramentas de versionamento de código Git ⁵ e GitHub ⁶. Mesmo com a experiência prévia adquirida durante a graduação, foi possível se aprofundar mais no uso dessas ferramentas.

Como parte prática dessa capacitação e visando uma gestão mais eficiente do desenvolvimento de código dos projetos de software exigidos na capacitação do estágio, cada estagiário foi responsável por criar e manter o seu próprio repositório de código. Assim, os monitores do estágio acessavam o que foi produzido e a prestavam a assistência necessária, em caso de dificuldades (Figura 6).

Figura 6 – Organização do Repositório Git



Fonte: próprio autor.

⁵ Disponível em: <https://git-scm.com>. Acesso em: 15 jan. 2024.

⁶ Disponível em: <https://github.com>. Acesso em: 15 jan. 2024.

Durante o treinamento com o repositório foram abordadas as atividades fundamentais de manejo, incluindo a criação e clonagem de repositórios, a realização de *commits* eficientes, o gerenciamento de *branches* e de conflitos e a colaboração em projetos compartilhados. Foi exigido que o repositório fosse organizado em *sprints*, cada uma em pastas específicas. Dentro de cada pasta foram armazenados os códigos correspondentes aos entregáveis de cada iteração, proporcionando uma documentação clara e segmentada de cada etapa de capacitação do estágio. Isso auxiliou os estagiários no acompanhamento de mudanças e na revisão de códigos específicos de cada *sprint* do estágio.

Inclusive, os projetos de *sprints* anteriores acabaram servindo como referência para os projetos seguintes, sendo bastante valioso para se revisar aprendizados já realizados, na medida em que a capacitação avançava.

Durante o estágio, foram realizadas 10 *sprints*, cada uma registrando os seus entregáveis (código e relatórios), que totalizaram 51 *commits* no Github.

3.2.3 Aprofundamento no Sistema Operacional Linux

O Linux é um sistema operacional bastante utilizado nas tarefas diárias de um engenheiro de dados, pois oferece um ambiente de execução robusto e versátil, com uma poderosa interface de linha de comandos para gerenciar processos, sistema de arquivos, entre outros. Em resumo, o Linux capacita os engenheiros de dados com as ferramentas e a flexibilidade necessárias para otimizar as tarefas de gerenciamento e processamento de dados, melhorando assim a produtividade e a eficiência no campo (FAHMY, 2023).

Em um primeiro momento do estágio, foi exigido o uso de uma infraestrutura computacional virtualizada, em que o sistema operacional Ubuntu Desktop precisou ser instalado em uma máquina virtual (VM) de um computador pessoal. O gerenciador de VMs utilizado foi o VirtualBox⁷.

Inicialmente nesse ambiente foram realizadas práticas com os comandos de terminal e com os atalhos do Ubuntu para a manipulação de arquivos e pastas, gerenciamento de processos em execução e gerenciamento de pacotes de instalação

⁷ Disponível em: <https://www.virtualbox.org>. Acesso em: 5 fev. 2024.

e de atualizações do sistema. Comandos como *cd*, *pwd*, *ls*, *kill*, *ps*, *rm*, *mkdir*, *mv*, *apt-get*, por exemplo, foram bastante explorados.

De posse desse conhecimento básico do sistema Ubuntu, o estágio passou a designar atividades envolvendo a utilização da ferramenta de computação em nuvem Elastic Compute Cloud (“EC2”), da Amazon Web Services (AWS⁸.) O “EC2” é um serviço de infraestrutura em nuvem categorizado como *Infrastructure as a Service* (IaaS), já que possibilita o provisionamento e o gerenciamento de instâncias de servidores virtuais na nuvem, conferindo flexibilidade, escalabilidade e controle sobre os recursos computacionais.

Nesse sentido, um IaaS é um modelo de computação em nuvem que fornece infraestrutura de TI sob demanda pela Internet que é usado por provedores de nuvem que fornecem recursos de hardware virtualizados para armazenamento, redes, servidores e outras infraestruturas de computação, podendo acessar esses recursos conforme necessário, pagando apenas pelo que usam, em vez de comprar e manter seu próprio hardware físico (AMAZON, 2023a).

Ao longo do estágio, foram adquiridos conhecimentos sobre as práticas de provisionamento, configuração e gerenciamento de instâncias “EC2”, bem como a realização de diferentes cenários, como o de criação de máquinas virtuais, configuração de redes e armazenamento e implantação de aplicações. As práticas com o “EC2” foram necessárias para o aprendizado de outras ferramentas da própria AWS e de outras tecnologias de análise de dados, na sequência do estágio.

3.2.4 Processo de manipulação de dados

O processo de Extração, Transformação e Carregamento (ETL) é essencial para integrar e preparar dados brutos de diversas fontes de formatos diferentes, automatizando a sua coleta, transformação e organização em um repositório central ou *data warehouse*. Regulado por regras de negócios, o ETL torna os dados aptos para análises avançadas, armazenamento eficiente e a aplicação de técnicas de *machine learning* (ML), proporcionando *insights* valiosos para *business intelligence*. (AMAZON, 2023b).

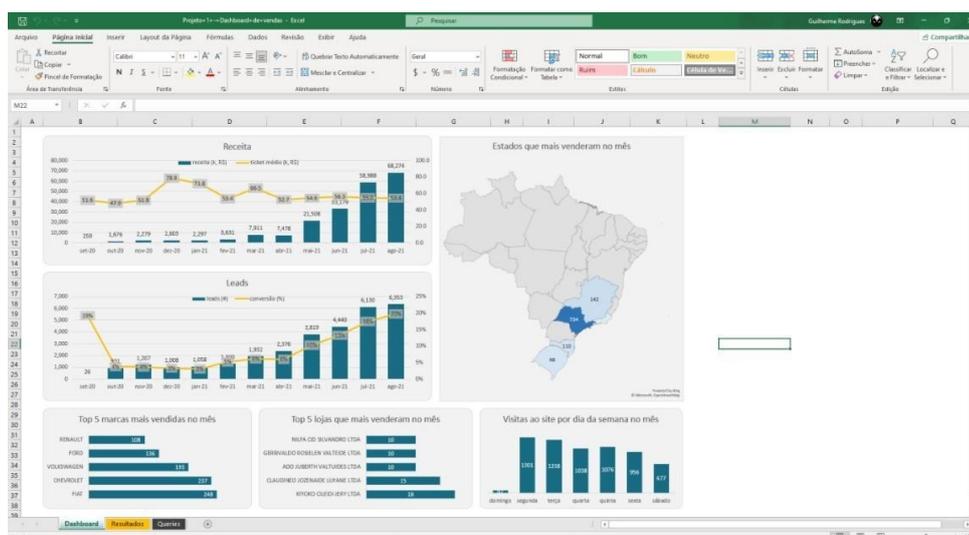
⁸ Disponível em: <https://aws.amazon.com/pt/> Acesso em: 5 fev. 2024 .

Durante o estágio, complementando esses conceitos, foram realizadas atividades práticas de SQL, envolvendo a elaboração de *queries* direcionadas a uma base de dados disponibilizada pela empresa em um formato de arquivo bastante comum, o.csv (*Comma-Separated Values*). Isso possibilitou explorar e extrair informações específicas, aplicando-se os conceitos de ETL de forma menos teórica e mais prática. Essa experiência proporcionou uma abordagem “mão na massa” na manipulação de dados do mundo real, melhorando as habilidades no desenvolvimento ETL em ambientes de *business intelligence*.

Em uma das *sprints* foi requisitado o desenvolvimento sob um painel informativo (*dashboard*) já pronto, por meio da aplicação de *queries* em SQL para a extração e transformação durante o processo de ETL, com o auxílio da ferramenta de administração de bancos de dados DBeaver⁹.

O *dashboard* ofereceu uma visualização gráfica resumida de informações relevantes sobre vendas de veículos, a partir da carga do arquivo .csv resultante do processo ETL. Conceitos e possibilidades na área de *business intelligence* foram explorados a partir dessa solução, considerando como dados brutos de formatos e fontes diferentes e, todavia, não interconectados, poderiam gerar *insights* valiosos em organizações, contribuindo para a tomada de suas decisões gerenciais e estratégicas (Figura 7).

Figura 7 – Dashboard Carregado com Dados Pós-Processados em ETL no Excel



Fonte: próprio autor.

⁹ Disponível em: <https://dbeaver.io> Acesso em: 31 mar. 2024 .

3.2.5 Exploração de conceitos de *Big Data*

O estágio ofereceu uma capacitação abrangente em *Big Data*, que explorou os fundamentos associados a essa área em ascensão na engenharia de dados e em constante evolução. Esse conceito refere-se ao gerenciamento de grandes volumes de dados que não podem ser facilmente tratados com os métodos tradicionais de armazenamento e processamento de dados.

Ele é caracterizado pelos "três Vs": volume (grande quantidade de dados), variedade (diversidade de fontes e formatos dos dados) e velocidade (necessidade de processamento rápido, às vezes em tempo real). Portanto, as organizações enfrentam desafios ao lidar com *Big Data*, mas também podem obter benefícios significativos, como redução de custos e ganhos de eficiência. As tecnologias de *Big Data* incluem coleta, armazenamento, processamento e análise de dados, com o objetivo final de obter insights valiosos para tomada de decisões (AMAZON, 2023c).

No início da jornada de aprendizado em *Big Data*, houve uma imersão em conceitos fundamentais, incluindo arquiteturas de processamento distribuído, tecnologias de armazenamento escalável e ferramentas analíticas especializadas. Essa exploração teórica estabeleceu a base necessária para compreender não apenas o potencial transformador do *Big Data*, mas também os desafios inerentes à sua gestão eficiente.

Apesar da abordagem teórica generalista, durante a realização do curso sobre *Big Data*, é importante destacar que não houve atividades práticas específicas nesse campo durante esse período. A ênfase se deu, meramente, no entendimento conceitual da área, o que limitou a experimentação direta desses conhecimentos em situações do mundo real. Entretanto, a compreensão teórica representou um alicerce valioso na continuidade do aprendizado e na exploração de tecnologias e oportunidades de aplicação relacionadas a essa área.

Conhecimentos como *data lake*, *data storage* e *dataware house* foram apresentados e diferenciados, assim como capacidades e utilidades de ferramentas da AWS, como o "Amazon S3" (*Simple Storage Service*), para criar *data lakes* personalizados para análise; e, o "Amazon Redshift" que usa o SQL (*Structured Query Language*) para analisar dados estruturados e semiestruturados em *dataware houses*.

3.2.6 Aperfeiçoamento na linguagem Python

A linguagem de programação Python é conhecida por sua versatilidade e eficiência. Durante uma das *sprints* do estágio, uma das tarefas de experimentação exigiu a manipulação de arquivos ETL, empregando-se essa linguagem. Aplicou-se o conceito de abstração de orientação a objetos (OO) para representar as entidades das fontes de dados distintas do experimento, tanto as de entrada (*input*), quanto as de saída (*output*).

Python apresentou-se bastante versátil nesse sentido, sobretudo pela disponibilidade de APIs para a manipulação de arquivos de diversos formatos que uma aplicação ETL costuma exigir. Por exemplo, a Pandas¹⁰ permite gerar e organizar bidimensionalmente os *dataframes* (como em planilhas).

Ainda, há a API PySpark para acesso ao Apache Spark¹¹, inclusive a partir de cargas de dados com Pandas (APACHE, 2024). O PySpark permite o uso de comandos SQL com operações de bibliotecas nativas em Python a partir da criação de *dataframes*, aplicando-se funções para análise de dados sobre conjuntos de dados resilientes distribuídos (em inglês, RDDs) geridos pelo Apache Spark. Esse último é capaz de processar dados usando alta concorrência, diretamente na memória (em *cache*), evitando-se à latência de leitura no acesso em disco (AMAZON, 2023d). No Spark há módulos para diversas tarefas de processamento de dados, *machine learning* e de grafos.

Já a API da biblioteca NumPy¹² foi essencial para facilitar a resolução de cálculos matemáticos complexos no processamento desses dados, fornecendo suporte para uso de *arrays* multidimensionais (vetores e matrizes) e uma variedade de funções para operações matemáticas avançadas.

Na implementação da lógica das etapas ETL em Python, foi empregada a programação estruturada, utilizando bibliotecas como *functools* e recursos como as funções *map*, *filter*, *reduce* e *lambda*. A função *map* é usada para aplicar uma função a todos os itens em uma entrada iterável. Já a função *filter* cria um iterador capaz de separar elementos que passarem em um teste específico. Já a função *reduce* agrega ou acumula valores de uma sequência em um único resultado. Por fim, as funções

¹⁰ Disponível em: <https://pandas.pydata.org>. Acesso em: 3 abr. 2024 .

¹¹ Disponível em: <https://spark.apache.org>. Acesso em: 3 abr. 2024 .

¹² Disponível em: <https://numpy.org>. Acesso em: 3 abr. 2024 .

lambda permitem a criação de funções anônimas para conveniência (redução de código), já que podem ser reusadas e repassadas facilmente como argumentos de entrada em chamadas de outras funções, como *map*, *filter* e *reduce*.

3.2.7 Implementação de Ambiente Docker e Uso do Apache Spark

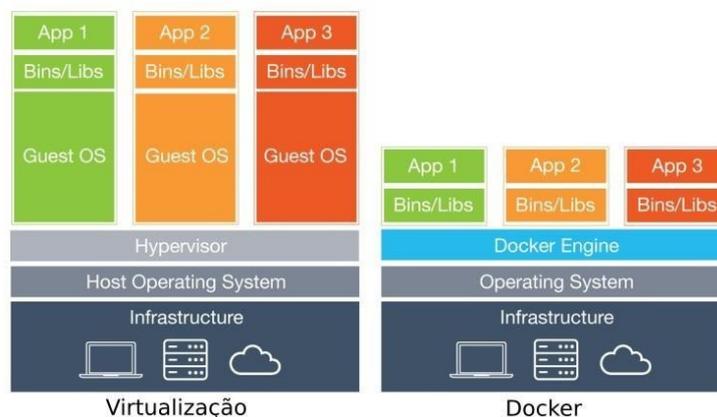
A implementação de ambientes de execução de aplicações é bastante facilitada pela plataforma Docker¹³. Ela proporciona a disponibilização de ambientes de execução isolados (contêineres), com características consistentes e específicas, a depender de cada tipo de aplicação a ser disposta, visto que um contêiner Docker encapsula no sistema operacional (SO) as dependências e configurações necessárias de cada tipo de aplicação (ABDALLA, 2018). Isso reduz o esforço comum relacionado com dependências de software e conflitos em ambientes de execução, eventualmente compartilhados por aplicações de natureza diferente, sob um mesmo SO.

O Docker como tecnologia de virtualização de processos também possibilita a portabilidade de seus contêineres, o que inclui a sua configuração e execução consistente, independentemente das configurações e do sistema de destino. Portanto, aplicações implantadas de forma integrada ao Docker simplificam essa etapa do desenvolvimento, oferecendo uma maneira eficiente de empacotar e distribuir aplicações, contribuindo para a replicação consistente e livre de conflitos em diferentes cenários.

Portanto, o Docker se diferencia da virtualização tradicional (que virtualiza todo o SO convidado sobre um SO hospedeiro), pois na verdade cria espaços isolados sob o mesmo SO, considerando versões diferentes de ambiente de execução de programas e de bibliotecas (Figura 8).

¹³ Disponível em: <https://www.docker.com>. Acesso em: 3 abr. 2024.

Figura 8 - Diferença do Docker e a Virtualização Tradicional



Fonte: (ABDALLA, 2018)

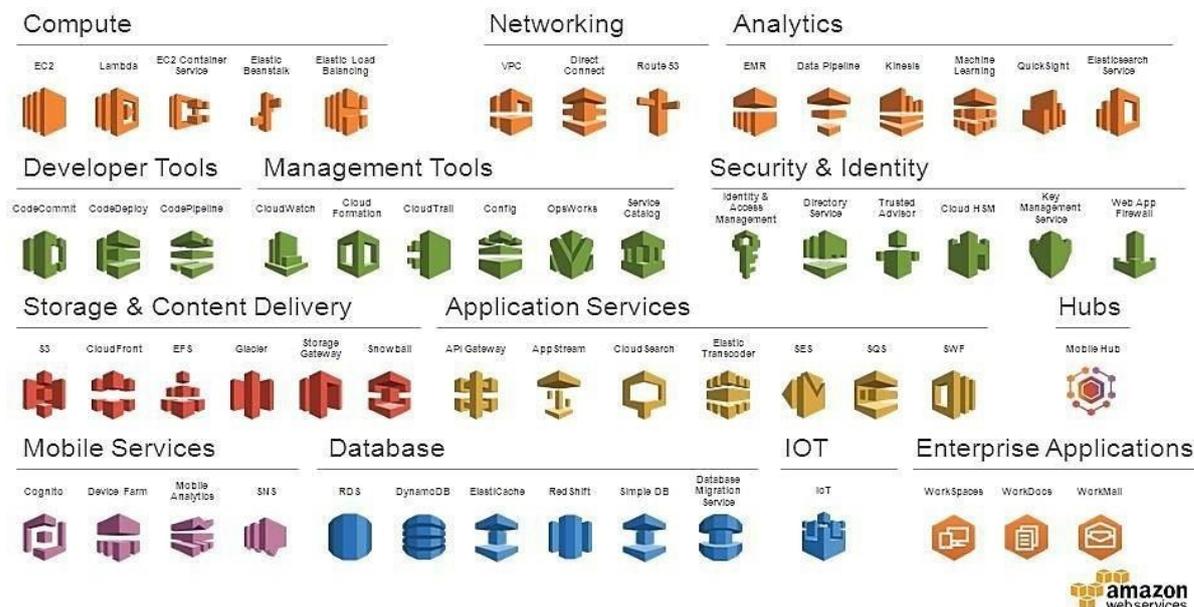
O Docker é bastante útil na implantação e no gerenciamento de *clusters* Apache Spark, que se torna mais simplificada e eficiente, bastando a criação de uma imagem Docker contendo todas as configurações e dependências necessárias, incluindo o próprio Spark instalado e configurado, bem como das demais bibliotecas necessárias, garantindo-se uma consistência na configuração de cada um dos nós de um *cluster* Spark adicional que precisar ser implantado.

Portanto, quando há a necessidade de dimensionar o *cluster* para lidar com uma carga de trabalho maior, pode-se iniciar mais contêineres em novos nós, agregado escalabilidade facilitada, com menos esforço.

3.2.8 Utilização dos serviços AWS

A Amazon Web Services (AWS) é uma plataforma líder em serviços em nuvem que oferece uma extensa gama de soluções para atender às demandas de empresas e desenvolvedores na era da computação em nuvem (Figura 9).

Figura 9 – Serviços da Amazon Web Services (AWS)



Fonte: (METRÓPOLE DIGITAL, 2024).

Lançada pela Amazon em 2006, a AWS se estabeleceu como uma referência no fornecimento de serviços de nuvem do tipo Infraestrutura como Serviço (em inglês, IaaS), de Plataforma como Serviço (PaaS) e de Software como Serviço (SaaS).

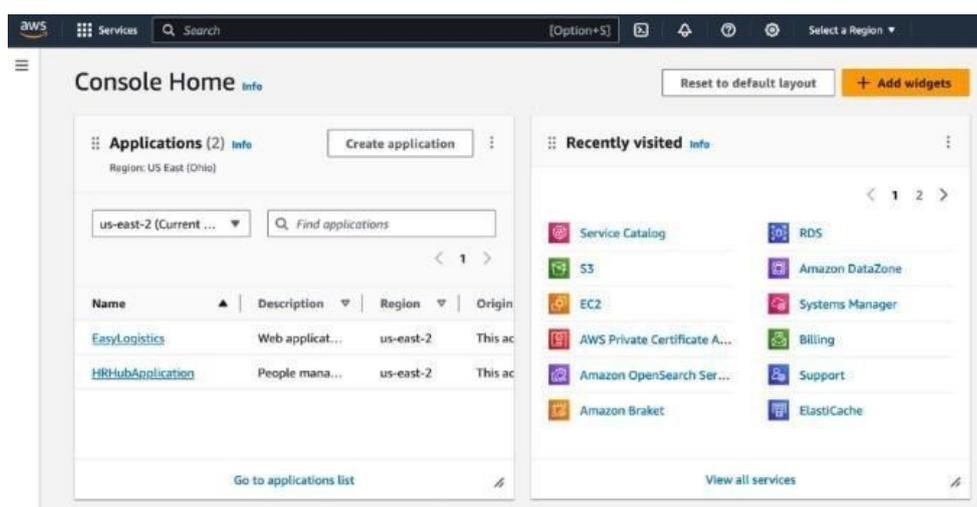
Com uma presença global, a AWS oferece escalabilidade, flexibilidade e confiabilidade, permitindo que organizações inovem, cresçam e gerenciem aplicativos em um ambiente altamente seguro e eficiente. Seus serviços abrangem desde computação em nuvem, armazenamento e banco de dados até inteligência artificial, análise de dados e Internet das Coisas (IoT), consolidando-se como uma ferramenta fundamental para diversas aplicações e setores (AMAZON, 2023a).

No estágio foram explorados diversos serviços da AWS para otimizar o processamento e a visualização de dados. O serviço “Amazon Athena”, por exemplo, foi empregado para consultas interativas diretamente nos dados armazenados no “Amazon S3”, proporcionando uma abordagem eficaz para análises. Já o “AWS Glue” auxiliou na preparação e transformação de dados, automatizando o processo de ETL (Extração, Transformação e Carregamento). Destaca-se que o “Amazon S3” serve como um repositório central escalável para armazenamento de dados.

Além disso, é notável ressaltar a exploração do serviço “AWS Lambda”, que não apenas possibilitou a execução de código sem a exigência de provisionamento

ou administração de servidores, mas, também, se destacou pela sua capacidade de ser automaticamente adaptável. Essa característica torna mais fácil a implementação de funções *serverless* de processamento de eventos em tempo real. A Figura 10 ilustra a console de serviços AWS, considerando o provisionamento de aplicações em nuvem.

Figura 10 – Console da Amazon Web Services (AWS)



Fonte: (YUN, 2023).

Dessa forma, foi possível verificar a importância da AWS como uma plataforma poderosamente confiável para implementar soluções de engenharia de dados, destacando sua capacidade de oferecer serviços especializados que atendem às necessidades específicas de processamento, armazenamento e análise de dados, em um ambiente adaptativo e dinâmico para disponibilização de aplicações.

3.2.9 Análise de Dados, Modelagem para Normalização, Dimensionamento e Multidimensionalização dos Dados

A modelagem de dados é uma atividade essencial para os analistas de dados. Uma das atividades do estágio exigiu a utilização de uma fonte de dados denominada TMDb¹⁴ (The Movie Database), que é uma base de dados mantida em comunidade,

¹⁴ Disponível em: <https://developer.themoviedb.org/reference/intro/getting-started>. Acesso em: 31 mar. 2024.

que cataloga filmes e programas de TV, a qual disponibilizada uma API REST de consumo.

Na atividade de estágio, os dados se concentraram em filmes do gênero romance. Portanto, essa base serviu como estudo de caso para se aprender sobre como normalizar, dimensionar e multidimensionalizar massas de dados para análise.

Na normalização é almejada uma organização eficiente dos dados, realizada por meio da estruturação de tabelas relacionadas, em que foi utilizado o ambiente “AWS Athena”. Essa abordagem não apenas tratou redundâncias nos dados, mas também as inconsistências.

Já no dimensionamento, objetiva-se reajustar os valores de campos para que correspondam a tipos, formatação ou escalas de valores específicos. Essa técnica foi aplicada com êxito em campos como os de avaliações e de popularidade de filmes, visando uniformizá-los para prevenção de distorções em análises subsequentes, também com o auxílio do “AWS Athena”.

Por fim, na multidimensionalização de dados, uma dimensão refere-se a uma categoria ou atributo pelos quais os dados podem ser organizados e analisados, como um gênero de filme, um diretor, um elenco ou um ano de lançamento. A implementação desses conceitos no “AWS” Athena facilita análises mais profundas, embora não se tenha envolvido uma indexação direta dos dados. Ao se criar dimensões sobre avaliações de usuários, datas de lançamento ou elenco de filmes, por exemplo, tornou-se possível segmentar e analisar os dados de forma mais granular, identificando tendências ao longo do tempo, preferências do público e padrões de desempenho.

Os serviços “AWS Glue” e o “AWS Lambda” desempenharam um papel fundamental no processo de ETL sobre a API REST TMDB, assegurando a velocidade e a integridade necessárias para análises mais aprofundadas.

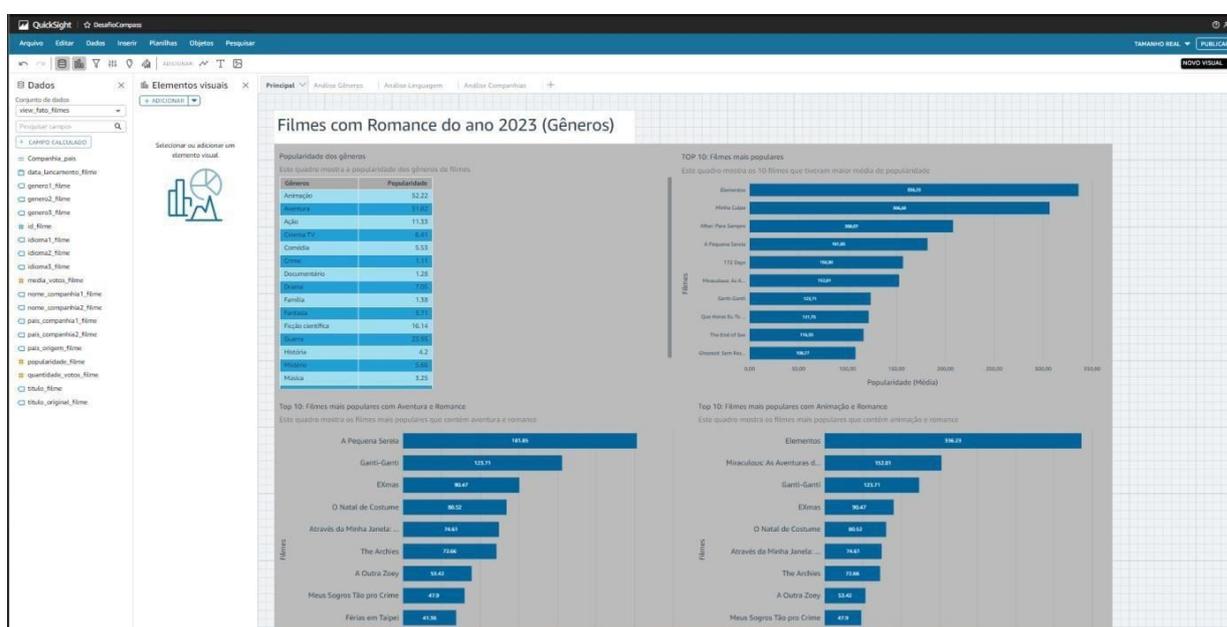
A análise resultante proporcionou *insights* importantes sobre o gênero romance, contribuindo para o desenvolvimento de estratégias mais informadas e embasadas em dados, cujos resultados foram armazenados no serviço “Amazon S3”.

A análise foi conduzida utilizando o “AWS Athena”, que ofereceu uma ferramenta poderosa para consultas SQL, diretamente nos dados armazenados no “Amazon S3”. Essa abordagem possibilitou uma exploração flexível e eficaz dos dados retirados do TMDB, utilizando Python para a sua extração e SQL para criação

de filtros. O conjunto de dados consistia em 92 páginas, cada uma contendo informações de 100 filmes, permitindo a aplicação dos filtros necessários para a criação das tabelas aplicadas sobre o “AWS Quicksight”.

Na etapa final, os resultados foram visualizados por meio do “AWS Quicksight” (Figura 11), que é uma ferramenta de *business intelligence* na nuvem. Essa plataforma possibilitou a criação de dashboards interativos, facilitando a compreensão e comunicação dos *insights* derivados da análise.

Figura 11 – Dashboard TMDb com AWS Quicksight



Fonte: próprio autor.

4 CONSIDERAÇÕES FINAIS

Após cinco meses de estágio foi possível alcançar avanços em diversas áreas da Engenharia de Dados. Os objetivos estabelecidos no início foram alcançados, evidenciando a compreensão e aprimoramento das habilidades essenciais nesse tipo de trabalho.

Além das expectativas acadêmicas, o estágio proporcionou uma oportunidade única de experimentar conceitos e práticas novos, além dos já conhecidos no curso superior de tecnologia, a exemplo da melhoria de habilidades: (i) no manejo de ferramentas como Git e GitHub no desenvolvimento colaborativo em equipe; (ii) no uso do ambiente Linux, com a aplicação de comandos e o gerenciamento de arquivos em terminal; (iii) na manipulação de dados por meio do SQL, considerando a elaboração de consultas complexas e a realização de operações ETL de maneira eficiente; (iv) análise de dados em Big Data com Apache Spark em contêineres Docker, considerando o tratamento de massas de dados via programação em Python, aliada ao uso avançado de bibliotecas como Pandas, Numpy e PySpark; (v) a aquisição e experiência na configuração de aplicações dessa natureza com serviços em nuvem da AWS, a partir do contato com ferramentas como “Athena”, “Glue”, “S3” e “Lambda”.

Apesar dos desafios enfrentados ao longo do estágio, como a curva de aprendizado em novas tecnologias e a adaptação a diferentes ambientes de trabalho, cada obstáculo foi encarado como uma oportunidade de crescimento e aprendizado.

A colaboração com a equipe foi essencial para superar esses desafios, tendo sido bastante proveitoso o trabalho com profissionais experientes e dispostos a compartilhar conhecimentos e soluções.

Uma sugestão para aprimorar ainda mais o estágio seria considerar a extensão do período de experiência, permitindo uma imersão mais profunda nas atividades e projetos, com mais estudos de caso (assim como ocorrido com o TMDb), possibilitando um desenvolvimento ainda mais substancial das habilidades e conhecimentos.

Em síntese, o estágio foi uma experiência extremamente valiosa que contribuiu significativamente para o desenvolvimento profissional, como também possibilitou a contratação no estúdio de *Data & AI*, após os quatro meses do término do estágio.

REFERÊNCIAS

ABDALLA, G. **Aula 1: Docker e Microsserviços - Introdução - documentação Docker Unleashed 1.5**. 2018. Disponível em: <https://docker-unleashed.readthedocs.io/aula1.html>. Acesso em: 3 abr. 2024.

AMAZON. **O que é IaaS? – Explicação sobre infraestrutura como serviço – AWS**. 2023a. Disponível em: <https://aws.amazon.com/pt/what-is/iaas/#:~:text=A%20IaaS%20é%20um%20modelo,nuvem%20acessíveis%20para%20pequenas%20empresas>. Acesso em: 1 abr. 2024.

AMAZON. **O que é ETL? – Explicação sobre extrair, transformar e carregar – AWS**. 2023b. Disponível em: <https://aws.amazon.com/pt/what-is/etl/#:~:text=O%20processo%20de%20ETL%20automatiza,até%20mesmo%20durante%20uma%20execução>. Acesso em: 19 jan. 2024.

AMAZON. **O que é Big Data? - Explicação sobre análise de big data - AWS**. 2023c. Disponível em: <https://aws.amazon.com/pt/what-is/big-data/>. Acesso em: 1 abr. 2024.

AMAZON. **O que é o Spark? – O que é o Spark? - Introdução ao Apache Spark e análises - AWS**. 2023d. Disponível em: <https://aws.amazon.com/pt/what-is/apache-spark/#:~:text=no%20Apache%20Hadoop,-,Como%20funciona%20o%20Apache%20Spark%3F,com%20a%20toler%C3%A2ncia%20a%20falhas>. Acesso em: 1 mar. 2024.

AMAZON. **Serviços de Computação em Nuvem - Amazon Web Services (AWS)**. 2024. Disponível em: <https://aws.amazon.com/>. Acesso em: 22 jan. 2024.

APACHE. **PySpark Overview – PySpark master documentation**. 24 fev. 2024. Disponível em: <https://spark.apache.org/docs/latest/api/python/index.html>. Acesso em 1 mar. 2024.

BELMONTE, P. Data Science Data Engineering: Explorando a Engenharia de Dados. **Blog da Awari**. 26 jun. 2023. Disponível em: https://awari.com.br/data-science-data-engineering-explorando-a-engenharia-de-dados-em-projetos-de-data-science/?utm_source=blog&utm_campaign=projeto+blog&utm_medium=Data%20Science%20Data%20Engineering:%20Explorando%20a%20Engenharia%20de%20Dados. Acesso em: 10 out. 2023.

COMPASS UOL. **Blog da Compass UOL - Compass é a nova marca adotada pela Compasso em fase de expansão global**. 8 nov. 2021. Disponível em: <https://blog.compass.uol/noticias/compass-e-a-nova-marca-adotada-pela-compasso-em-fase-de-expansao-global/>. Acesso em: 17 jan. 2024.

COMPASS UOL. **Home | Compass UOL**. 2024a. Disponível em: <https://compass.uol/en/home/>. Acesso em: 15 jan. 2024.

COMPASS UOL. **O que fazemos | Compass UOL**. 2024b. Disponível em: <https://compass.uol/pt/o-que-fazemos/>. Acesso em: 17 jan. 2024.

COMPASS UOL. **Sobre nós | Compass UOL**. 2024c. Disponível em: <https://compass.uol/pt/sobre-nos/>. Acesso em: 17 jan. 2024.

COMPASS UOL. **Sicredi D2E | Compass UOL**. 2024d. Disponível em: <https://compass.uol/pt/cases/sicredi-d2e/>. Acesso em: 17 jan. 2024.

FAHMY, M. A. **Important 4 Data Engineering Skills that don't get the proper hype**. 16 set. 2023. Disponível em: <https://medium.com/art-of-data-engineering/important-4-data-engineering-skills-that-dont-get-the-proper-hype-cc4f3281a4bf#:~:text=With%20Linux,%20data%20engineers%20can,editing%20code%20and%20data%20files>. Acesso em: 8 fev. 2024.

IBM, **What is software development?** Disponível em: <https://www.ibm.com/topics/software-development>. Acesso em: 10 out. 2023.

MARTINS, L. Desafios de gerenciamento de projetos de software e como lidar com eles. **IT Forum**. 14 fev. 2023. Disponível em: <https://itforum.com.br/noticias/desafios-de-gerenciamento-de-projetos-de-software-e-como-lidar-com-eles/>. Acesso em: 10 out. 2023.

METRÓPOLE DIGITAL. **Material Didático - IMD**. Disponível em: <https://materialpublic.imd.ufrn.br/curso/disciplina/4/67/9/7>. Acesso em: 4 abr. 2024.

MICROSOFT. **O que é controle de versão? - Azure DevOps**. 5 out. 2023. Disponível em: <https://learn.microsoft.com/pt-br/devops/develop/git/what-is-version-control>. Acesso em: 18 jan. 2024.

MIND MASTER TREINAMENTOS. **Scrum: A Metodologia Ágil Explicada de forma Definitiva**. Disponível em: <https://mindmaster.com.br/scrum/>. Acesso em: 18 jan. 2024.

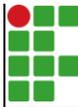
NUNES, J. **O que faz um engenheiro de dados? Profissão tem salário alto e deve bombar em 2024**. G1. 24 nov. 23. Disponível em: <https://g1.globo.com/tecnologia/noticia/2023/11/24/o-que-faz-um-engenheiro-de-dados-profissao-tem-salario-alto-e-deve-bombar-em-2024.ghtml>. Acesso em: 4 jan. 2024.

SCHWABER, K.; SUTHERLAND, J. **O Guia do Scrum – O Guia Definitivo para o Scrum: As Regras do Jogo**. Nov. 2020. Disponível em: <https://scrumguides.org/docs/scrumguide/v2020/2020-Scrum-Guide-PortugueseBR-3.0.pdf>. Acesso em: 4 abr. 2024.

UNIFOR. Qual a diferença entre analista, cientista e engenheiro de dados? - Blog Melhor Profissão - UNIFOR. **Blog Melhor Profissão**. 06 set. 2023. Disponível em: <https://unifor.br/web/melhor-profissao/qual-a-diferenca-entre-analista-cientista-e-engenheiro-de-dados>. Acesso em: 10 out. 2023.

YUN, C. New my Applications in the AWS Management Console simplifies managing your application resources | Amazon Web Services. **AWS News Blog**. 30 nov. 2023. Disponível em: <https://aws.amazon.com/pt/blogs/aws/new-myapplications-in-the->

[aws-management-console-simplifies-managing-your-application-resources/](#). Acesso em: 22 jan. 2024.

	INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DA PARAÍBA
	Campus Monteiro - Código INEP: 25284940
	Pb-264, S/N, Serrote, CEP 58500-000, Monteiro (PB)
	CNPJ: 10.783.898/0008-41 - Telefone: (83) 3351-3700

Documento Digitalizado Ostensivo (Público)

Relatório de estágio

Assunto:	Relatório de estágio
Assinado por:	Guilherme Cosa
Tipo do Documento:	Relatório
Situação:	Finalizado
Nível de Acesso:	Ostensivo (Público)
Tipo do Conferência:	Cópia Simples

Documento assinado eletronicamente por:

- **Guilherme Rodrigues da Costa, ALUNO (202015020014) DE TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS - MONTEIRO**, em 18/03/2025 22:53:13.

Este documento foi armazenado no SUAP em 18/03/2025. Para comprovar sua integridade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifpb.edu.br/verificar-documento-externo/> e forneça os dados abaixo:

Código Verificador: 1424801
Código de Autenticação: 1d09269f7a

