



**UNIVERSIDADE
FEDERAL RURAL
DE PERNAMBUCO**



Desenvolvimento de uma plataforma de compartilhamento e análise de Dados com ênfase em LGPD

**Relatório Técnico relativo ao Trabalho de Conclusão Curso
do Bacharelado em Sistemas de Informação na modalidade Empresa**

Aluno

Heitor Augusto Gomes Monteiro

Orientador

Victor Wanderley Costa de Medeiros
Departamento de Estatística e Informática (DEINFO)

12 de maio de 2023

Heitor Augusto Gomes Monteiro

Desenvolvimento de uma plataforma de compartilhamento e análise de dados com ênfase em LGPD

Relatório Técnico apresentado ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Universidade Federal Rural de Pernambuco – UFRPE

Departamento de Estatística e Informática

Curso de Bacharelado em Sistemas de Informação

Orientador: Victor Wanderley Costa de Medeiros

Recife

Março de 2023

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

- M775d Monteiro, Heitor Augusto Gomes
Desenvolvimento de uma plataforma de compartilhamento e análise de Dados com ênfase em LGPD /
Heitor Augusto Gomes Monteiro. - 2023.
20 f. : il.
- Orientador: Victor Wanderley Costa de Medeiros.
Inclui referências.
- Trabalho de Conclusão de Curso (Graduação) - Universidade Federal Rural de Pernambuco,
Bacharelado em Sistemas da Informação, Recife, 2023.
1. Setor Público. 2. Governança de dados. 3. LGPD. I. Medeiros, Victor Wanderley Costa de, orient. II.
Título

Heitor Augusto Gomes Monteiro

Desenvolvimento de uma plataforma de compartilhamento e análise de dados com ênfase em LGPD

Relatório Técnico apresentado ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Aprovado em 27 de Abril de 2023.

BANCA EXAMINADORA

Victor Wanderley Costa de Medeiros (Orientador)
Departamento de Estatística e Informática
Universidade Federal Rural de Pernambuco

Gabriel Alves de Albuquerque Júnior
Departamento de Estatística e Informática
Universidade Federal Rural de Pernambuco

Agradecimentos

Em primeiro lugar, eu gostaria de aproveitar esta oportunidade para agradecer a minha mãe. A pessoa mais importante da minha vida, que sempre esteve presente me apoiando. Graças aos seus esforços eu tive uma boa educação, tive a oportunidade de ingressar numa universidade pública e concluir o curso de computação que tanto almejava. Não consigo resumir em palavras meus sentimentos por você, Ana Flávia, mas sou extremamente grato por tudo, mainha. Também gostaria de agradecer a minha tia Fabiana, a meu tio Wanderlan e as minhas primas Larissa e Laura, pelo seu apoio, presença e torcida para que eu pudesse finalizar o curso. Gostaria de agradecer a Marisa, minha tia-avó, que mora comigo e minha mãe desde sempre, sempre sendo uma pessoa alegre que é e contagiando a todos com sua presença. Amo todos vocês.

Muito obrigado professor Victor, meu orientador, do Departamento de Estatística e Informática da Universidade Federal Rural de Pernambuco, que sempre foi atencioso e solícito. Sei que não tive a melhor das participações como seu orientando, tive bastante dificuldade de gerenciar meu tempo para desenvolver o trabalho e reportá-lo. Mas ainda assim, você sempre esteve presente para tudo o que desse. Agradeço a todos os ensinamentos que pude ter durante minha trajetória no curso e pelo seu apoio e atenção. Gostaria de também agradecer a toda equipe de docentes e profissionais presentes no curso de Sistemas de Informação, por todos os ensinamentos e experiências que obtive. Agradeço a todos.

Queria agradecer a Gabriel Mac'Hamilton, meu melhor amigo, que conheço desde o período da alfabetização e que hoje trabalhamos juntos na mesma empresa, a Fábrica de Negócio. Desde sempre me apoiando e progredindo comigo, e graças a ele, que eu tive a oportunidade de ingressar no mercado de trabalho e aprender coisas novas. Queria agradecer a todos os amigos que pude fazer na empresa, Hamilton, Mailson, Paulo, Ewerton, Agrinaldo, Ericks, Italo Y., Felipe, Iago, Lucas Y., Lucas A. e Pedro M., que me acolheram, apoiaram e me moldaram no profissional que sou hoje. Queria agradecer mais uma vez a Mailson, pelo apoio na construção deste trabalho, me guiando e orientando com informações internas da empresa. Muito obrigado.

Gostaria de lembrar dos amigos que pude fazer em toda minha trajetória, que estão comigo desde o ensino fundamental Arthur, Fábio, Germano, Gabriel F., Huberto, Pedro R., Rafael e Vinícius. E também dos amigos que pude fazer durante o curso, Ebony, Eliana, Eric, Gabriel L., Kimbelly, Leandro, Lucas, Luciano, Matheus R., Matheus P., Nicolás, Nicollas, Ulysses, além de outras pessoas que, de alguma forma contribuíram para minha jornada de aprendizagem cooperativa e desafiadora. Muito obrigado.

Posso dizer que fui extremamente sortudo de ter tido oportunidade de conhecer e conviver com pessoas maravilhosas, que participaram de todo o meu processo de crescimento e que continuam me apoiando. Eu sou profundamente grato a todos!

Resumo

O desenvolvimento tecnológico destacou a importância dos dados como um ativo estratégico na administração pública e a necessidade de os governos desenvolverem estruturas de governança de dados e mecanismos de entrega relacionados à infraestrutura e padrões de dados. Compartilhar informações é útil para o planejamento estratégico. Melhorar a previsão e a resposta às necessidades do usuário e fornecer melhores serviços e políticas públicas. No entanto, no setor público, muitas vezes você pode encontrar muitas informações contidas em vários bancos de dados. Requer desenvolvimento empresarial e soluções de troca de dados acessíveis. O uso adequado da tecnologia da informação e comunicação (TIC) pode reduzir ineficiências nos processos governamentais, permitir o compartilhamento de arquivos e dados entre órgãos governamentais, eliminar erros e reduzir o tempo necessário para as transações. No entanto, o compartilhamento de dados entre órgãos governamentais muitas vezes não é padronizado, resultando em dificuldades na adoção de diferentes processos, falta de conhecimento holístico dos dados corporativos, indicadores de consumo insuficientes e falta de mecanismos de *feedback* do usuário. Neste trabalho, será apresentada minha atuação como desenvolvedor da Fábrica de Negócio para a resolução de um problema de compartilhamento de dados no setor público, mais especificamente, na construção de um portal que opera como um orquestrador dos serviços presentes na solução, que hoje é utilizada por órgãos do estado no seu ambiente de produção.

Sumário

1	Introdução	1
2	A empresa e sua atuação	2
3	Desenvolvimento realizado na empresa	3
3.1	A problemática e a solução proposta	3
3.2	Tecnologias utilizadas	8
3.3	Contribuição	9
4	Dificuldades encontradas	9
5	Impactos da sua formação no seu trabalho	10
6	Conclusão	14

1 Introdução

Com a velocidade do desenvolvimento tecnológico, o mercado e a economia mundial vem comprovando a prevalência dos dados sobre os demais ativos organizacionais. E na administração pública não seria diferente: um setor público orientado a dados os reconhece como um ativo estratégico na concepção e entrega de políticas públicas e serviços [1].

Na busca pelo refinamento de processos de disponibilização de dados e pela criação de uma cultura de tomada de decisão com base em evidências, governos precisam desenvolver estruturas sólidas de governança de dados e mecanismos de entrega relacionados a infraestruturas de dados e padrões. Afinal, o compartilhamento de dados pode beneficiar no planejamento estratégico, melhorando a antecipação e resposta às necessidades dos usuários, fornecendo melhores serviços e políticas públicas que promovam integração, acesso, compartilhamento e uso de dados em todo o setor público à luz da legislação vigente [2].

No setor público é comum encontrar uma abundância de dados dispostos em bases distintas que precisam ser compartilhadas entre diversos órgãos. Além disso, é desejável que esses órgãos estejam inseridos nas boas práticas de *compliance* e, assim, consigam apresentar uma boa governança desse processo de compartilhamento; é também natural em iniciativas públicas a crescente dependência de novas tecnologias – como, por exemplo, a que se convencionou chamar *Big Data* – que lidem bem com o rápido crescimento da quantidade e variedade de dados e fontes digitais, que precisam ser coletados, armazenados, recuperados e processados com alta velocidade e segurança.

Não é incomum os governos atenderem às demandas de compartilhamento de forma singular e sem padrão; com papéis, processos, tecnologias e formalizações diferentes a depender dos órgãos envolvidos. Isso levava ao desperdício de recursos, retrabalho das pessoas na operacionalização das demandas e fragilidade na segurança, contribuindo para o surgimento da necessidade de fazer convergir as demandas de compartilhamento para uma solução corporativa e acessível que permitisse o armazenamento e processamento veloz de grandes volumes de dados digitais em formatos variados.

Diante desse cenário, foi desenvolvida uma estratégia de compartilhamento de dados, que priorizasse o reuso de dados, pessoas, processos e tecnologias, devidamente amparada por arranjos institucionais e regras formalizados em decreto, e implementada na Plataforma de Compartilhamento e Análise de Dados, os quais serão apresentados neste trabalho.

Garg, Agarwal e Sherry [3] afirmam que do ponto de vista da eficiência, para bom uso dos dados é melhor capturar os dados apenas uma vez e armazená-los e compartilhá-los para permitir a integração dos usuários entre os diversos interessados. Um único local de compartilhamento pode permitir o acesso a serviços, consultas e transações relacionadas. Esse compartilhamento de dados é fundamental para modernizar o governo, facilitando a entrega eletrônica de serviços governamentais, promovendo um atendimento de qualidade ao cliente, reduzindo a duplicação e erros. Mas esses dados abrangem um volume muito grande de espaço e para processá-los é fundamental o uso de ferramentas de alto desempenho, que possam se beneficiar de computação paralela e distribuída para acelerar o processamento e prover respostas rápidas.

A aplicação adequada das Tecnologias da Informação e Comunicação (TIC) pode possivelmente

reduzir o número de ineficiências nos processos ao permitir o compartilhamento de arquivos e dados entre os departamentos governamentais, contribuindo, assim, para a eliminação de erros de procedimentos manuais, reduzindo o tempo necessário para as transações [4].

No âmbito do governamental, é comum o compartilhamento de dados entre os órgãos não dispor de uma estruturação padronizada, o que leva a diversas dificuldades, como: (1) a adoção de processos de formalização e de padrões tecnológicos variados para obtenção de acesso, transmissão e consumo dos dados entre órgãos; (2) o desconhecimento holístico e transparente sobre os dados e informações corporativos gerados no âmbito do governo, eventualmente disponíveis para consumo pelos demais órgãos; (3) a insuficiência de métricas de consumo e mecanismos de *feedback* de usuários que permitissem aferir a qualidade dos dados existentes; e (4) a falta de conhecimento e controle dos compartilhamentos existentes entre os órgãos da administração pública estadual, assim como entre esses órgãos e instituições externas ao governo, sobretudo após a promulgação da Lei Geral de Proteção de Dados Pessoais (LGPD).

Com essas dificuldades latentes relacionadas ao mundo governamental, foi pensado e levantado a necessidade de criação de um ambiente de compartilhamento de dados, cuja estratégia e processo de desenvolvimento forneceria (1) Autonomia aos encarregados dos dados nas atividades de gestão e compartilhamento; (2) Capacidade de lidar com grande volume de dados; (3) Segurança e Privacidade dos dados pessoais.

2 A empresa e sua atuação

A Fábrica de Negócio é uma *Startup GovTech* pernambucana, que atua em diversas áreas do setor público, oferecendo soluções para diversas secretarias e órgãos de esferas municipais, estaduais e federais; que desejam soluções inovadoras para obtenção de informações no poder de decisão e agilidade nos processos de produção. Sua sede está localizada no bairro da Madalena, na cidade do Recife. A *Startup* foi fundada em 2016, pelos sócios Hamilton Alves e Mailson Santos, que estrearam a iniciativa com um projeto que ajudou a secretaria de saúde do município de Recife a identificar pontos de foco de arboviroses, relatados ao órgão, auxiliando à equipe de tomada de decisão a identificar os locais e mobilizar alguns profissionais da saúde para irem até esses locais e averiguar a situação.

Atualmente, a empresa possui produtos focados em governo nas áreas de Governança de Dados (*Angus GDL*), Inteligência Fiscal (Plataforma *Intelektas*) e Inteligência Investigativa (*Xerlock*). Além disso, através de processos licitatórios oferecem serviços de desenvolvimento de soluções de *analytics*. Já foram atendidos diversos clientes do setor privado, dentre eles: Jornal do Comércio; Ferreira Costa e Porto de *Suape*; focando hoje grandes esforços no setor público, que entre os principais clientes estão: Empresa Municipal de Informática (EMPREL); Secretarias de Saúde de Recife; Secretarias de Finanças de Recife; Controladoria Geral do município do Recife; Agência de Tecnologia da Informação do Governo do Estado de Pernambuco (ATI-PE); Secretaria da Controladoria Geral do Estado de Pernambuco (CGE-PE); Processamento de Dados do Amapá (PRODAP); etc.

Durante todos os seus anos de atuação, a Fábrica de Negócio atuou em diversos projetos no mundo governamental e atualmente está focado em governança de dados e em plataformas investigativas. Foi uma das primeiras empresas do Brasil a poder utilizar o selo *GovTech* (através de um

processo de aceleração realizado em SP em 2019), indicando ser uma empresa que produz produtos de alto impacto para o governo. Além de ter figurado no *100 Open StartUps* dentre as Top 10 *Govtech*, nos anos de 2021 e 2022.

A empresa atualmente divide os projetos em duas grandes áreas voltadas para dados: a área de engenharia de dados, onde profissionais visam modelagem de dados e o melhor ajuste do dado, desempenho e escala para o projeto em questão e a equipe de *Front-End*, que trabalha de modo *Full Stack Web*, utilizando toda uma pilha de tecnologias voltadas para o desenvolvimento de produtos de dados.

Ingressei na empresa como estagiário, em setembro de 2018 e finalizei meu período de estágio após 1 ano, em 2019, quando passei a me tornar um colaborador definitivo. Hoje, 4 anos depois, sou líder técnico da área de *Front-End*, desenvolvendo produtos de dados como também apoiando tecnicamente outros desenvolvedores. As equipes de engenharia de dados e de *Front-End* contam hoje com uma equipe de 5 pessoas, cada, desempenhando papéis fundamentais na entrega de soluções de valor ao cliente e posicionados em diferentes produtos que constituem o arcabouço da empresa.

Durante meu histórico como integrante da equipe de *Front-End*, pude experienciar e desempenhar diversas atividades que englobam o desenvolvimento de produtos de software, que vão desde a idealização do produto em projetos de design até a produção de documentações do produto finalizado. Seguindo uma linha lógica de desenvolvimento sequencial de um produto de software, ocupei tarefas de desenvolvimento de *wireframes* e protótipos; construção e manutenção de códigos de software; documentação de manuais de usuário e diagramas de BPMN (*Business Process Model and Notation*); reuniões e comunicação com clientes acerca de assuntos técnicos; configurar ambientes de infra-estrutura para execução de projetos em produção; e pesquisas e estudos para evolução das ferramentas de trabalho para a área de desenvolvimento de software. Hoje, minha atuação foca principalmente na parte de desenvolvimento de software.

3 Desenvolvimento realizado na empresa

Esta seção visa detalhar a problemática do projeto e solução proposta, além das tecnologias utilizadas e como se integram bem como a contribuição real do projeto para o governo.

3.1 A problemática e a solução proposta

Durante o decorrer do projeto, foram feitos estudos voltados para o entendimento do contexto de dados do estado, como funcionava o processo de compartilhamento de dados e suas regras, como esses dados eram armazenados, quais eram os principais sistemas que movimentam grandes volumes de dados e como seria preciso abordar tecnologias que suprimisse o desafio de armazenar, processar e disponibilizar o dado com segurança. O cenário encontrado, conforme a Figura 1, contava com diversas entidades, representadas por órgãos públicos do governo, que possuem sistemas transacionais diferentes e tinham a necessidade de mais controle, autonomia e independência para realizar a carga, gestão e compartilhamento desses dados. Além disso, um fator determinante para

o sucesso do projeto era que a plataforma tivesse a capacidade de garantir o uso dos principais conceitos de LGPD e garantir o compartilhamento seguro do dado.

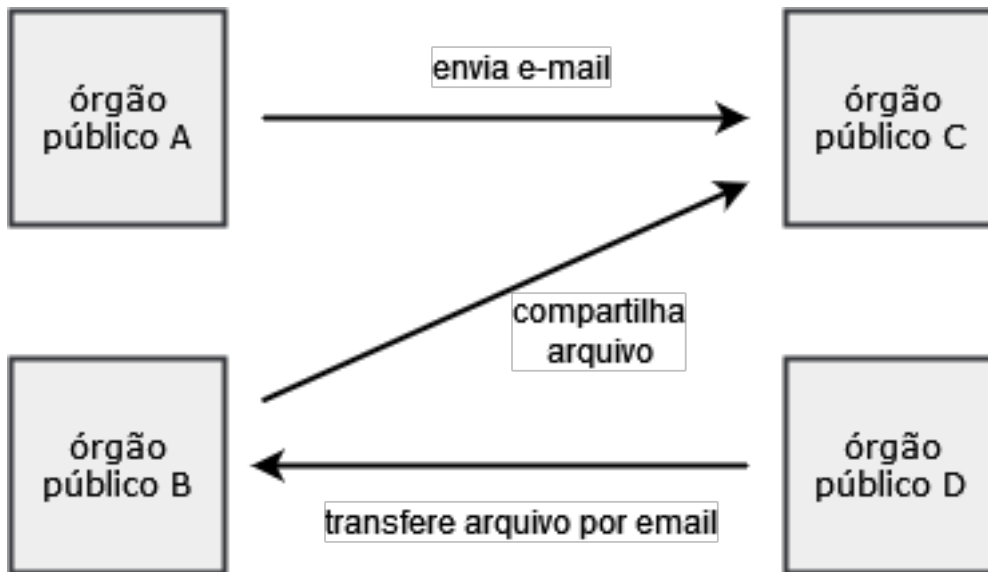


Figura 1: Cenário encontrado de compartilhamento dos dados

A solução que deveria ser proposta tinha como premissas de ser baseada em tecnologias *open-source*, expansível, segura e que tivesse a capacidade de ser executada no hardware disponível no data center do governo. Depois de estudarmos algumas opções, a equipe do projeto chegou a algumas situações:

- Precisaríamos criar uma estrutura modular para o projeto, pois seria o ideal para trabalhar com o nível de complexidade exigido pelo projeto, permitindo que diferentes grupos de desenvolvimento trabalhem de forma isolada em suas áreas específicas. Um grupo seria responsável pela camada de dados, outro pela infraestrutura e montagem dos *scripts* dos servidores, e um terceiro grupo seria voltado para o *front-end* da aplicação.
- Seria fundamental uma interface de compartilhamento de dados que fosse transparente para a maioria dos sistemas e desenvolvedores, possibilitando o acesso aos dados.
- O nível de segurança que precisaríamos, dado o nível de confidencialidade dos dados, não pode se restringir apenas ao acesso físico à máquina.
- Os papéis relacionados aos dados e aos metadados devem estar em conformidade com a LGPD, garantindo a visibilidade do dado dentro do aspecto principal da lei.
- Precisaríamos ter um segundo fator de autenticação para garantir usuários apenas dentro da rede do governo.

Então foi pensado em uma arquitetura modular baseada em microsserviços, onde teríamos “atores” responsáveis por módulos do projeto, conforme apresentado na Figura 2 abaixo:

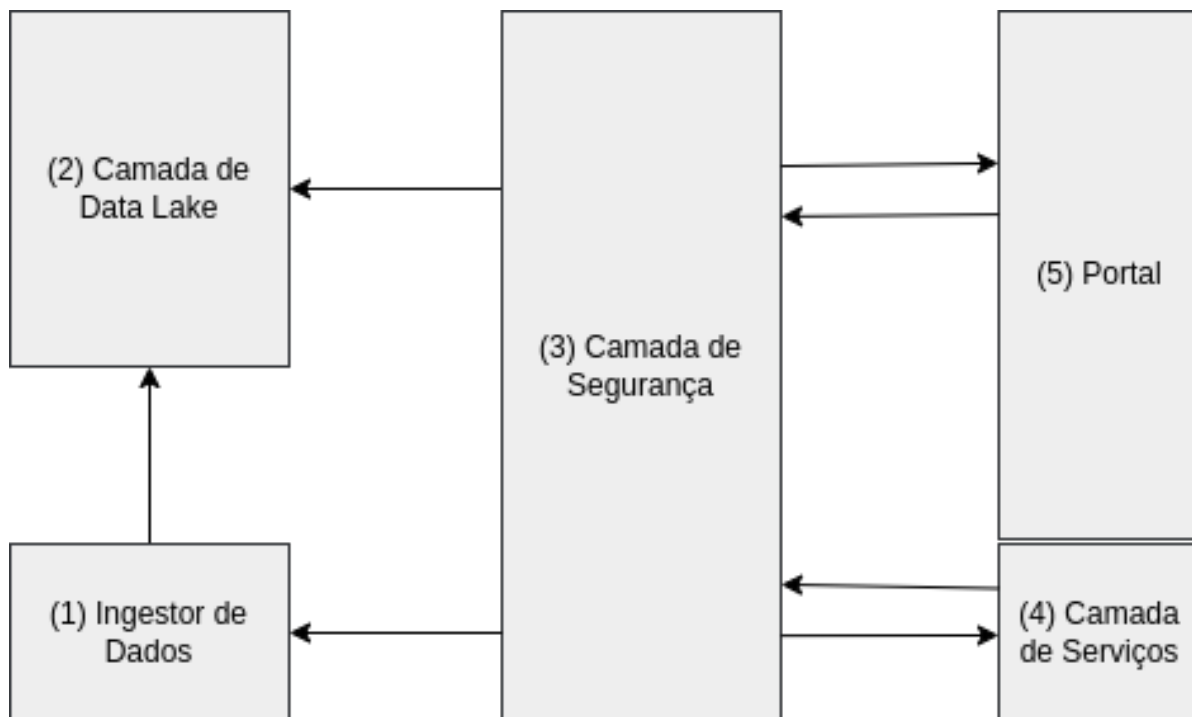


Figura 2: Arquitetura modular do projeto – Atores

1. Ingestor de Dados - Responsável pelo processo de carga de dados e estruturação do dado dentro da camada de *Data Lake* do projeto.
2. Camada de *Data Lake* - Camada composta por tecnologias *open-source* para armazenamento de grande volumes de dados e de forma não estruturada.
3. Camada de Segurança - Seu papel é garantir que as máquinas estejam fechadas na rede e que só fossem acessíveis com *ticket* de segurança e estritamente ao dado necessário.
4. Camada de Serviços - Conjunto de *API REST* para consumo do dado, contendo conjunto de regras para validação do usuário em relação ao dado que foi permitido o acesso. Esse método de compartilhamento foi escolhido devido ao processo ser largamente utilizado nos principais sistemas e serviços que o governo utiliza.
5. Portal - Camada de Aplicação, um grande portal gerenciador dos outros “atores” e onde basicamente o usuário vai interagir com as configurações do ambiente, carga dos dados, gestão de acesso e papéis relacionados aos dados.

Dessa forma, teríamos equipes diferentes trabalhando em seus módulos específicos, dando mais robustez ao projeto e garantindo que as equipes trabalhassem focadas nos seus módulos ao invés de pensar na arquitetura como um todo e no fim eles seriam ligados como serviços disponíveis entre si. Todos os módulos trabalham com o conceito de *API REST* para comunicação, conforme representado pelas setas da Figura 2, as quais indicam o sentido da dependência para relacionamentos entre ambientes distribuídos em uma rede de diversas máquinas.

Como camada de regras principais relacionadas à solução proposta, haviam regras relacionadas à LGPD (dado sensível, cadastral ou comum) além de perfis de usuários com características de

acesso diferentes. O portal em si funcionaria como um grande orquestrador dos serviços que os outros módulos deveriam executar, conforme Figura 3.

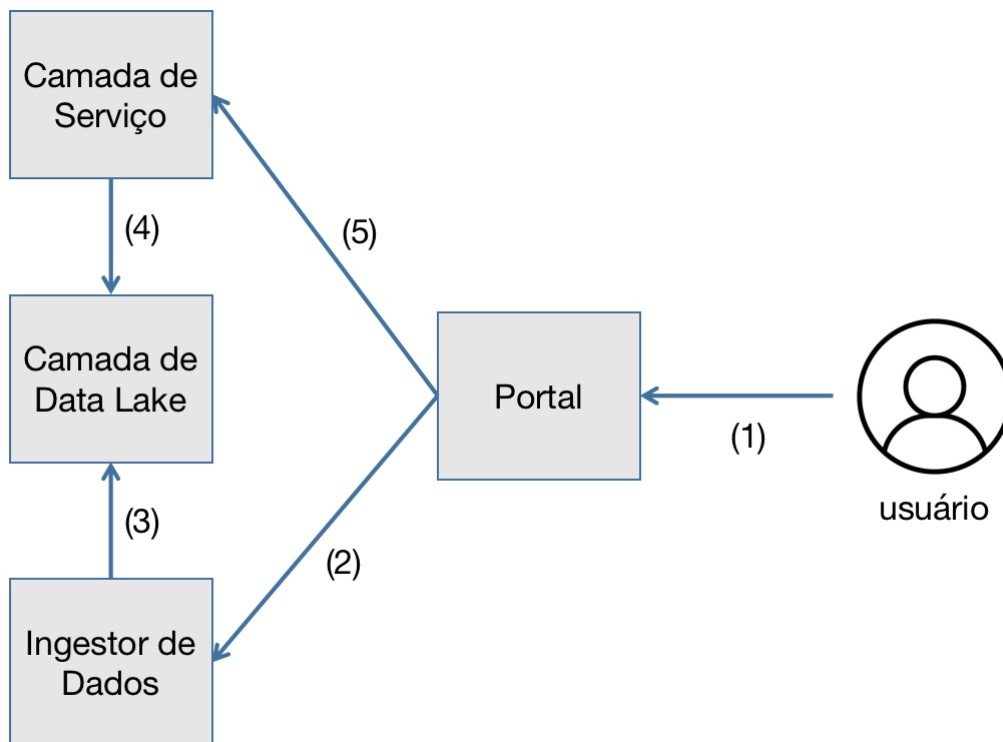


Figura 3: Visão de serviços – Gestão e consulta dos dados

- (1) - Usuário interage com o portal, podendo requisitar acesso a conjuntos, criar conjuntos e configurar processo de ingestão.
- (2) - Evento de disparo do portal para o ingestor, indicando que deve começar a realizar o processo de ingestão do conjunto x, o *scheduler* de ingestão fica sob controle do portal.
- (3) - Processo de carga dos dados processados para armazenamento dentro da estrutura do *data lake*.
- (4) - Processo de busca de dado presente no Data Lake, para entrega ao usuário, em formato de *JSON* via requisição de protocolo *HTTP*.
- (5) - Evento realizado pela camada *API REST* onde o portal indica o que o usuário tem acesso e em que condições de acesso.

Nesta visão, a camada de segurança não teve o mesmo foco como na visão de arquitetura. Isto ocorre, pois entende-se que cada módulo se comunica por esta camada como fator de segurança, que está presente em toda a plataforma. Assim é possível ter uma percepção maior sobre qual a finalidade dos demais módulos ao trabalharem um cenário simples de gestão e consulta dos dados.

Um dos pontos críticos do projeto é a capacidade que o portal terá que ter para gerenciar todos esses eventos e armazenar cada passo em *logs* para posterior validação, caso necessário. Como

os outros módulos ficam apenas “aguardando” serem requisitados, o controle dessa fila de eventos precisa ser robusto para atender às necessidades do projeto.

O módulo pelo qual fui responsável, (5) Portal, passou por um processo de validação após o levantamento inicial. Foi necessário modelar um protótipo para criar a identidade visual da plataforma e apresentar ao cliente uma representação do ambiente antes do desenvolvimento efetivo, evitando retrabalhos devido a desentendimentos. Após várias discussões sobre telas, usabilidade e visibilidade, as funcionalidades que estariam presentes no portal foram validadas, de acordo com os requisitos descritos pelo cliente.

Durante esse período, ocorreram diversas reuniões para correções e adições, que culminaram na aprovação do estado inicial do portal e no início do processo de desenvolvimento dessa camada. Uma das regras importantes para o projeto foi a definição de papéis relacionados à LGPD, destacando três papéis:

- Administrador - Responsável por gerenciar a plataforma e suas características.
- Controlador - Papel relacionado à LGPD, atuando como “dono do dado”. Tem a responsabilidade de gerenciar a camada de dados da plataforma, podendo criar e gerenciar conjuntos de dados, definir seus metadados e características, e conceder permissões de acesso. É o principal ponto de interação com o dado e o usuário final, pois é responsável pela ingestão do dado e disponibilização do mesmo.
- Usuário Final - Pode ver os conjuntos de dados que tem acesso e solicitar o acesso ao conjunto/dado.

No processo de construção do portal, foram desenvolvidas tanto a parte *front-end* quanto o *back-end*, que contaram com as seguintes características:

- *Front-end* - Ambiente onde estão todas as construções de telas e funcionalidades de interação com o usuário. Foi necessário garantir uma maior performance de renderização dos dados, utilizando técnicas de requisições sob demanda, em que os dados e páginas são carregados apenas quando o usuário acessa determinada parte do portal. Isso evitou requisições que trazem todos os dados em uma única requisição, congestionando a rede e o *back-end*, o que tornaria as operações mais lentas para a primeira requisição.
- *Back-end* - Possui todas as regras de negócio e funcionalidades essenciais para o pleno funcionamento do portal, como autenticação de segundo fator, comunicação com o banco de dados e execuções periódicas. Nesta parte, foi construída uma *API* específica para gestão de filas, que desempenha um papel importante na comunicação com a camada do ingestor, para dar início ao processamento de ingestão de dados e envio de e-mails de autenticação e alertas do portal. Além disso, foi adotada uma estratégia de performance para manipulação de dados de painéis no *back-end*, em que as estruturas de dados específicas para renderização de componentes eram tratadas diretamente em requisições de banco de dados e, posteriormente, combinadas em código.

Para além disso, o portal contou com a presença de diversas funcionalidades implementadas em ambos os ambientes, que exercem suas responsabilidades de forma colaborativa. As principais são:

- Autenticação de dois fatores - Login de usuário e senha, com camada de segurança para validação do código de verificação.
- Painel estatístico - Área presente para administradores, apresentando dados de monitoramento do *Data Lake*, com informações de consultas realizadas por outros usuários e ingestão dos dados.
- Conjuntos de dados - Acessado por administradores e controladores que realizam atividades básicas para cadastro e gerenciamento dos conjuntos de dados presentes no *Data Lake*. Esta é a funcionalidade principal, que constitui a base de atividades que são trabalhadas no ambiente e atende sobretudo o problema exposto. Nela estão presentes configurações com base nas regras da LGPD, onde os conjuntos são categorizados quanto ao seu nível de LGPD, classificação de privacidade, e sensibilidade dos dados. Possibilitando assim disponibilizar apenas os conteúdos considerados como público para usuários gerais e informações sensíveis para os usuários permitidos.
- Catálogo de dados - Presente a todos os usuários, ambiente para visibilidade e solicitação de acesso a conjuntos de dados disponíveis, que serão liberados de acordo com os critérios avaliados pelos Controladores.

3.2 Tecnologias utilizadas

A arquitetura foi então projetada para funcionar com módulos de alto desacoplamento se integrando por serviços, conforme a Figura 2 apresentada anteriormente.

Descrevendo a base tecnológica dos módulos, temos:

1. Ingestor de Dados - *Java* e *Pentaho Data Integration*
2. Camada de Data Lake - *Hadoop* e *MongoDB*
3. Camada de Segurança - *Kerberos* e *PrestoDB*
4. Camada de Serviços - *Python* e *Flask*
5. Portal - *Javascript*, *Nodejs*, *Reactjs*, *Redis* e *MongoDB*.

No módulo (5) Portal, do qual fui responsável pela construção, foram utilizadas algumas ferramentas que pudessem auxiliar no desenvolvimento:

- O protótipo foi construído utilizando o *Adobe XD*, que serviu como plataforma de inspiração e prototipação interativa do projeto.
- No *front-end* foram utilizadas bibliotecas como o *Redux*, para gerenciamento de estados compartilhados e *Redux-Thunk*, como middleware de abstração de lógica do *Redux*; e *React-Router*, para gerenciamento de rotas.
- Para a camada de *UI* e *UX* foram escolhidos o *Echarts*, para construção de painéis gráficos interativos; *Material-UI*, como ferramenta de componentes interativos, responsivos e funcionais que aplicam o padrão do *Material Design*.

- No *back-end* foram utilizadas bibliotecas como o *framework Express*, que traz consigo ferramentas que facilitam o desenvolvimento de *APIs* com o *Nodejs*.
- *Mongoose*, *ODM* responsável pela comunicação com o banco de dados principal do portal, o *MongoDB*; possuindo utilitários de validação, esquematização e consulta das coleções de dados.
- *Bull*, um gestor de filas que integra diretamente com o banco de dados *Redis*, utilizado para o processamento dos serviços de envio de e-mail e início do evento de ingestão dos dados.
- E não menos importante o *nodemailer*, que permite a integração com serviço de envio de e-mails.

Essas ferramentas em conjunto foram a base tecnológica para a produção do portal, e cada uma representou bem seu papel na arquitetura que montamos para a camada de visualização e regras do portal.

3.3 Contribuição

O processo de construção do portal iniciou em abril de 2020 e foi ao ar em 2021, sendo divulgado por veículos oficiais do estado. Neste momento, a plataforma encontra-se em pleno funcionamento e recebendo novas atualizações constantemente, há cerca de 2 anos. Atualmente, há a presença de alguns órgãos do estado que fazem uso da plataforma de compartilhamento e análise de dados, e outros que estão aos poucos acessando e fazendo uso do ambiente. Como evoluções mais marcantes, pode-se destacar um novo processo de ingestão onde foi necessário expandir a capacidade de abstração da camada de serviço no portal, dando mais opções de ingestão, e interagindo diretamente com a equipe de engenharia de dados do projeto; e evoluções no processo de compartilhamento do dado derivado do uso e de necessidades dos clientes da plataforma, onde foram desenvolvidos filtros e formas de redução dos dados para apoiar no compartilhamento entre níveis de acesso.

A partir desse momento, ficou ainda mais claro a importância que esse projeto traria para o estado, e quão importante foi para a empresa, que agora possuía um novo projeto original capaz de atender às necessidades de outros estados. Além disso, esse também foi um período de grande crescimento profissional, já que houve novos aprendizados técnicos e pessoais, um reconhecimento maior de minha posição como colaborador e maiores responsabilidades para com a empresa, tendo em vista que novos desafios foram se tornando mais frequentes e minha presença se tornando cada vez mais necessária.

4 Dificuldades encontradas

Por grande parte do tempo fui o único desenvolvedor do portal e portanto, muitas funcionalidades tive que realizar com certa rapidez para evitar atrasos nas entregas das *sprints* semanais. Apesar da urgência de entregar os pontos nas datas acordadas, também precisei me atentar às questões de segurança; correções de *bugs*; e ajustes pedidos pelo cliente, conforme o trabalho avançava.

Nos momentos em que tive a presença de outro desenvolvedor comigo, eu precisei apoiá-lo nas dificuldades que lhe ocorressem, e seguimos a estratégia de separação de atividades, onde cada qual desenvolvia partes complementares do projeto, ou seja, um trabalhava na parte *front-end* enquanto o outro ficava com a parte *back-end*.

Durante o processo, enfrentei novas necessidades de busca de conhecimento e aprendizado para resolver problemas. Muitas vezes, investi bastante tempo lendo documentações de bibliotecas, buscando exemplos e vídeos na internet para compreender o funcionamento e a teoria por trás de assuntos específicos. Essas situações me tornaram familiarizado com o processo de leitura de documentações e absorção de novas informações. Quando a documentação não era clara ou deixava dúvidas, precisei consultar o código fonte para entender como a ferramenta funcionava e como poderia integrá-la ao processo de trabalho.

Durante a etapa de homologação do portal, foram realizados diversos testes, muitos dos quais geraram comportamentos inesperados, levando à necessidade de interromper o trabalho atual para fazer correções imediatas (*hotfix*). Nesse processo de constante necessidade de retornar a um ponto anterior do projeto onde ocorria um erro específico, percebi a necessidade de estudar mais a fundo e implementar uma estratégia mínima de versionamento com o *Git*, para evitar ter que voltar manualmente o código para uma versão anterior e partir dali com a correção feita. Houve casos de perda de código durante esse processo, resultando em retrabalho, mas serviu como um importante momento de aprendizado para entender os cuidados necessários ao realizar ações de mudança de versão e a importância de aplicar uma estratégia bem definida para navegar entre os pontos de interesse.

5 Impactos da sua formação no seu trabalho

Durante minha formação no curso de graduação em Sistemas de Informação, tive a oportunidade de ter contato com uma grande quantidade de conhecimentos que foram fundamentais para minha evolução profissional. Vivenciei situações que me prepararam para o dia-a-dia no campo da tecnologia da informação e construí uma rede de contatos com pessoas talentosas. Os primeiros períodos do curso foram essenciais para minha evolução na área, pois me prepararam para a vivência prática das atividades que exerço em 80% do meu tempo. Algumas cadeiras tiveram um impacto considerável na minha atuação, tais como:

- **Introdução à programação:** Aprendi minha primeira linguagem de programação com Python. Tive contato com noções básicas de algoritmo e estruturas de dados. Além disso, tive experiências em *pair programming* na realização de projetos da disciplina, o que acabou levando a um aprendizado mútuo. Apesar de não trabalhar com Python atualmente, a transição para JavaScript foi tranquila, pois percebi semelhanças entre as linguagens que me auxiliaram no entendimento da sintaxe;
- **Matemática discreta:** Pude desenvolver noções básicas, e ao mesmo tempo, essenciais de lógica com princípios booleanos de tabela verdade, teoria dos conjuntos, indução, função, etc. Todas essas etapas me auxiliaram na construção de um pensamento lógico e que complementou em paralelo com a cadeira de Introdução à programação. Os projetos de aplicação de programação para resolver problemas de matemática discreta também foram muito importan-

tes para interligar os conteúdos. Na plataforma de compartilhamento e análise de dados, pude utilizar esse conhecimento para criar a lógica de papéis de usuários, onde cada papel integra as permissões dos papéis inferiores numa espécie de hierarquia, seguindo a ideia da teoria dos conjuntos;

- **Introdução à Teoria da Computação:** Resgatou um conteúdo que foi aprendido em matemática discreta, com a teoria dos conjuntos e evolui para a introdução às expressões regulares e operadores gramaticais, com apresentação de máquinas de estados. Teve um projeto muito importante de construção de uma máquina de Turing, para aplicação de operações matemáticas básicas de soma, subtração e divisão, que exigiu a aplicação prática de todo o conhecimento adquirido na cadeira. Essa experiência foi muito importante, pois traduzi na prática com código a teoria escrita no livro de apoio da cadeira;
- **Algoritmos e Estruturas de Dados:** Tive muitas experiências práticas aplicando os conhecimentos teóricos da disciplina, principalmente relacionados à otimização de código. Isso ocorreu graças às introduções às estruturas de dados, como listas, tabelas *hash* e árvores, e ao entendimento dos custos de trabalho de um algoritmo, que envolvem noções de complexidade. O projeto de construção de uma árvore vermelha e preta foi enriquecedor para entender a fonte de desenvolvimento das estruturas de dados e também para aplicar ensinamentos teóricos do livro base da cadeira. Um exemplo interessante foi quando enfrentei um problema de processamento de grandes quantidades de dados para apresentação em um gráfico. A consulta levava cerca de 25 minutos para retornar algum valor, mas isso ocorria devido ao alto custo de código. Tive um *insight* ao aplicar o conceito de tabelas *hash*, cuja consulta a uma chave tem custo constante - $O(1)$. Ao aplicar esse conhecimento, pude melhorar a consulta que antes levava 25 minutos para apenas 5 segundos. Esse feito confirmou para mim a importância de avaliar o uso correto de estruturas de dados para obter ganhos na otimização de códigos;
- **Modelagem de Programação Orientada a Objetos:** Tive minha primeira experiência com a linguagem de programação Java, que incluiu uma introdução ao paradigma de orientação a objetos. Aprendi os conceitos fundamentais, como classes, objetos, herança, polimorfismo e encapsulamento. Também fui introduzido aos diagramas UML, com foco em diagramas de classes, casos de uso e pacotes. Esses conhecimentos em orientação a objetos foram muito úteis em meu primeiro projeto com uma equipe maior, onde construímos um aplicativo mobile nativo em Android com Java. Pude aplicar os conceitos de orientação a objetos na organização e arquitetura do projeto, fazendo uso de classes e métodos encapsulados. Isso ajudou na estruturação do código e na separação de responsabilidades entre as diferentes partes do aplicativo;
- **Fundamentos de Engenharia de Software:** Tive minha primeira experiência trabalhando de forma individual no desenvolvimento de um projeto com o objetivo de solucionar um problema do mundo real, enquanto precisava reportar o andamento a um membro da equipe de gerência. Nesta disciplina, fui introduzido aos métodos ágeis, incluindo a aplicação do Scrum em conjunto com o Kanban. Durante o projeto, tive a oportunidade de aplicar os conceitos aprendidos na disciplina de Modelagem de Programação Orientada a Objeto na construção de documentos de requisitos, bem como na criação de diagramas. Essa experiência foi muito útil para desenvolver uma mentalidade de produção textual, habilidades de redação e planejamento na descrição e

documentação de projetos;

- **Fundamentos em Bancos de Dados e Projeto de Banco de Dados:** Ambas as disciplinas me introduziram ao mundo dos bancos de dados e aos sistemas gerenciadores de bancos de dados relacionais. Na disciplina de Fundamentos em Bancos de Dados, aprendi sobre os conceitos de modelagem de dados, incluindo os diagramas do modelo entidade-relacionamento (ER), a linguagem SQL, modelagem normal, restrições de integridade e operadores da álgebra relacional. Já na disciplina de Projeto de Banco de Dados, pude aplicar na prática os conceitos aprendidos na disciplina de Fundamentos em Bancos de Dados, construindo um projeto completo de banco de dados. Embora eu não trabalhe diariamente com bancos de dados relacionais, esse conhecimento foi essencial para entender e manipular dados armazenados em bancos de dados. Compreendi a importância da modelagem correta do banco de dados para garantir a eficiência e integridade dos dados. A modelagem com o banco de dados em MongoDB da plataforma construída, segue inspirações nas regras de normalização relacional para coleções de dados NoSQL de um banco orientado a documentos;
- **Processo de Desenvolvimento de Software e Gerência de Projetos de Software:** Duas cadeiras que andam juntas, porém com objetivos diferentes. Em FES, o foco é o trabalho em equipe para produzir uma solução de valor para resolver um problema real, enquanto alinha os próximos passos com um aluno no papel gerente. Em GPS, o foco é em ser o gerente responsável por guiar uma equipe de desenvolvedores à uma solução. Ambas as cadeiras simulam o dia-a-dia de um trabalho real em uma empresa de tecnologia. E dessas eu pude sentir a dificuldade que é estar no papel de um gerente com uma equipe desmotivada e sem cooperação, e tendo que fazer o possível para fazer a entrega dar certo. Ambas as cadeiras transpassaram experiências comuns de atuar em equipes de desenvolvimento de software e que foram úteis para vida profissional;
- **Projeto de Sistemas Distribuídos e Redes e Sistemas de Internet:** Ambas as disciplinas têm uma sólida base teórica que me proporcionou um melhor entendimento de como funcionam os protocolos de comunicação em redes. Em Redes e Sistemas de Internet, pude aprofundar meu conhecimento sobre os protocolos HTTP e HTTPS, bem como os protocolos de transporte TCP e UDP, e o protocolo de internet (IP), que são fundamentais para a comunicação na internet. Em Projeto de Sistemas Distribuídos, pude compreender os conceitos de *clusters* e servidores em nuvem, e aplicar esses conceitos em um projeto prático em conjunto com ambas as disciplinas. Essa experiência prática me permitiu entender e aplicar o conceito de processamento paralelo em várias máquinas executando *scripts* independentes para processar dados em tempo real e apresentá-los em um painel de dados agregados. Essa abordagem é similar ao conceito de microsserviços, onde diferentes partes de uma aplicação podem ser distribuídas em diferentes máquinas, cada uma executando processos individuais que compõem o todo. Esse conhecimento expandiu minha compreensão sobre como lidar com sistemas distribuídos e me permitiu utilizar ferramentas adequadas para esse tipo de tarefa. Atualmente, aplico esse conhecimento na separação de programas em diferentes máquinas, visando ganhos de desempenho e escalabilidade;
- **Linguagens de Programação:** Essa disciplina me proporcionou a oportunidade de experimentar diferentes paradigmas de programação, além da orientação a objetos, e compreender

seus usos em diferentes casos. Através do projeto realizado nesta disciplina, pude aprender o uso de ferramentas de ciência de dados, como o pandas em Python, para o processamento e análise de dados. Isso me permitiu compreender uma nova estrutura de dados voltada para o processamento de dados em alta performance e me livrar do pensamento de execuções manuais, substituindo por funções encapsuladas, resultando em um aumento significativo de desempenho e manutenibilidade do código. Essa disciplina também me conscientizou sobre a importância de aplicar boas práticas de programação em projetos, a fim de garantir a reusabilidade do código, facilitar a leitura e permitir a adição de novas funcionalidades de forma eficiente. Essa abordagem de programação limpa (*clean code*) tem sido aplicada em meu trabalho profissional, buscando sempre construir projetos com uma estrutura organizada e de fácil compreensão, minimizando a ocorrência de programas de baixa qualidade, que podem aumentar o custo de produção e manutenção;

- **Empreendedorismo e Interface Homem-Máquina:** As disciplinas de Empreendedorismo e Interface Homem-Máquina (IHM) foram complementares e me proporcionaram uma experiência valiosa em ambos os campos. A disciplina de Empreendedorismo foi fundamental para entender os processos de criação e desenvolvimento de ideias, semelhantes aos de uma incubadora ou aceleradora de *startups*. Aprendi a documentar projetos de forma eficiente, habilidade que tenho utilizado com frequência, inclusive na criação do manual do usuário para a plataforma que foi desenvolvida como parte do projeto. Na disciplina de IHM, pude experimentar técnicas de inovação e compreender melhor os conceitos de *User Interface* (UI) e *User Experience* (UX), e como aplicá-los para melhorar o design de produtos com base em *feedbacks* de terceiros. Essa experiência foi semelhante ao participar dos processos de incubação e aceleração do Inova UFRJ, onde pude aplicar o conhecimento prático adquirido na academia. Com muito esforço, eu e minha equipe conseguimos finalizar o processo com a criação de uma nova *startup*, chamada Skye Tecnologia, na qual participo como sócio;
- **Princípios de Software Básico e Segurança e Auditoria de Sistemas:** Nessas disciplinas, tive a oportunidade de vivenciar na prática e na teoria os processos de infraestrutura de software e configuração de sistemas operacionais. Compreendi a importância de manter a segurança para proteger contra acessos e execuções indevidas. Em Princípios de Software Básico, pude ter meu primeiro contato com a linguagem de programação C e seu uso na construção de *drivers* para controle de hardware. Em Segurança e Auditoria de Sistemas, compreendi casos de ataques conhecidos atualmente, assim como métodos para prevenir essas brechas e aplicar boas práticas de defesa e segurança. Através de atividades práticas de configuração de servidores, pude perceber que a segurança abrange várias áreas do conhecimento, tornando-se essencial para profissionais de todas as áreas evitarem falhas em suas esferas de atuação. Na plataforma construída, pude aplicar essas técnicas de defesa e monitoramento do sistema no ambiente de produção, aprimorando meus conhecimentos em software para prevenir invasões e ataques a terceiros, em nível de aplicação.

6 Conclusão

Durante minha atuação na Fábrica de Negócio, experimentei um considerável crescimento profissional. Aprendi e vivenciei situações que só poderia ter adquirido em ambientes de produção de soluções de impacto social. Tive contato com diversos pensamentos e conhecimentos que agregam muito à minha vida. Transitei de estagiário a líder técnico da equipe de Front-End, aumentando minha presença em participações e interações com os clientes. Além disso, evolui constantemente o nível das tecnologias utilizadas no desenvolvimento cotidiano. Todas essas experiências contribuíram para me tornar o profissional que sou hoje, sempre buscando evoluir e agregar valor em meu trabalho, e colocando em prática os novos aprendizados adquiridos. Tive a oportunidade de adquirir diversos conhecimentos nessa vasta área de Tecnologia da Informação, e aprimorei minhas habilidades interpessoais, estimuladas pelas cadeiras de administração, gestão e trabalho em equipe do curso.

Em todo meu histórico na Fábrica de Negócio, passei por diversos projetos bem sucedidos. Posso citar por exemplo, minha participação como desenvolvedor júnior em 2019, apoiando na construção de um projeto que nasceria de um processo de aceleração pela MPLabs, programa de inovação aberta do ministério público de Pernambuco, no desafio de “Soluções para Inteligência Investigativa”. O projeto foi batizado de Xerlock e foi vencedor nesse processo. Logo em seguida no começo do ano de 2020, durante a época da situação emergencial causada pela pandemia da Covid-19, participei como desenvolvedor júnior na construção da plataforma de BI do “Atende em casa”, uma realização da prefeitura do Recife e conjunto com o estado de Pernambuco, que teve como objetivo de reduzir o número de contrações do vírus ao realizar atendimentos médicos em casa pelo aplicativo. A plataforma de BI foi disponibilizada para a equipe de profissionais de saúde e utilizada para divulgar os informes de casos confirmados.

Após meu trabalho com o “Atende em casa”, participei do desenvolvimento da plataforma de compartilhamento e análise de dados com ênfase em LGPD, como desenvolvedor principal. Foi um projeto que contribuiu significativamente para a minha evolução como profissional e trata-se de um projeto de grande importância para o estado e de valor inestimável para a sociedade, uma vez que faz parte de um processo de transformação digital com o objetivo de centralizar os dados e facilitar o compartilhamento, reduzindo o desperdício de tempo e recursos, bem como a redundância e complexidade dos dados. Isso possibilita a elaboração de projetos conjuntos entre órgãos governamentais com menos obstáculos. Com base nisso, já existem planos futuros em desenvolvimento para alcançar a população em geral, com o objetivo de oferecer maior comodidade no acesso e visualização dos dados.

Em seguida, participei também da construção de ferramentas que compõem o que hoje é chamado de HUB da Saúde, da Prefeitura do Recife. Esse HUB possui vários portais de usos específicos, disponíveis apenas para os profissionais de saúde, como painéis de vacinômetros e indicadores de arboviroses. O projeto de indicadores de arboviroses foi, inclusive, o primeiro projeto que fundou a Fábrica de Negócio e passou por um processo de modernização de tecnologias e incorporação de novas funcionalidades.

Referências Bibliográficas

- [1] I. Peña-López *et al.*, “Digital government index: 2019 results,” 2020.
- [2] OECD, “The path to becoming a data-driven public sector, oecd digital government studies,” 2019.
- [3] K. M. Garg, N. Agarwal, and A. M. Sherry, “Maneuvering best business practices to improve the quality of e-governance services,” *Journal of Internet Banking and Commerce*, vol. 9, 2004.
- [4] V. Ndou, “E-government for developing countries: Opportunities and challenges.” *Electron. J. Inf. Syst. Dev. Ctries.*, vol. 18, no. 1, pp. 1–24, 2004.