Relatório final de atividades

Projeto: Patente. IA: estruturação de dados de patentes para sistema de busca baseado em Inteligência Artificial

Coordenador do Projeto:	**Washington Luís Ribeiro de Carvalho Segundo**		
Coordenação:	**Coordenação Geral de Informação Científica e Técnica – CGIC**		
Nome do(a) pesquisador(a):	**Rodrigo Gonçalves de Branco**		
Vigência da Bolsa de Pesquisa:	**06/07/2026**	**Até:**	**05/10/2026**
Encerramento em:	**05/10/2026**		
Ref. Fundação:	**32990**	**Ref. Ibict:**	**0580/2026-41**

Descrição das Atividades da Bolsa de Pesquisa:

1 - Estudo para definição e validação da arquitetura tecnológica. Meta 6- Implementação de novo sistema em ambiente de produção; 2 - Elaborar relatório técnico final de bolsa, consolidando resultados, atividades desenvolvidas e contribuições metodológicas da pesquisa.

1 INTRODUÇÃO

O presente relatório consolida as atividades desenvolvidas no âmbito do projeto "31849 - Serviços PI - INPI", cujo foco é a estruturação de dados de patentes para um sistema de busca baseado em Inteligência Artificial. A infraestrutura tecnológica da solução baseia-se em um fluxo de dados complexo, envolvendo a ingestão a partir de bases legadas do INPI (Informix e Oracle), processamento em um Datalake estruturado em camadas utilizando a Arquitetura Lakehouse Medalhão (Bronze, Prata e Ouro) - Anexo A, e disponibilização via APIs REST e painéis de consumo.

Durante o período de vigência desta bolsa, a atuação concentrou-se na Meta 6, referente à implementação do novo sistema em ambiente de produção. Embora a concepção arquitetônica já estivesse definida previamente, o foco do trabalho consistiu no refinamento, ajuste e estabilização da infraestrutura subjacente. Isso envolveu extensas discussões e implementações de orquestração com Docker Swarm e Portainer, pipelines de dados com Apache Airflow, Apache Spark, Logstash e Elasticsearch, além de aprimoramentos críticos na segurança (implementação de tokens JWT) e na organização do código-fonte através da segregação de repositórios.

2 MÉTODO

A metodologia adotada para a consolidação da arquitetura baseou-se em práticas de Engenharia de Plataforma e DevOps, visando garantir a resiliência, segurança e manutenibilidade do ambiente de produção. Os procedimentos foram divididos em três frentes principais de atuação:

Ajustes de Orquestração e Infraestrutura Base: A orquestração dos contêineres foi estruturada utilizando Docker Swarm, escolhido por sua aderência às necessidades do projeto e facilidade de gerenciamento através da interface do Portainer. Foram realizados ajustes finos no ambiente, que já se encontrava em operação, para

otimizar o consumo de recursos e garantir a alta disponibilidade dos serviços essenciais, como os nós do Elasticsearch e os workers do Airflow e Spark.

Segurança e Comunicação: Para assegurar a proteção dos dados trafegados e o controle de acesso às APIs REST do portal de serviços, foram conduzidas discussões e implementadas soluções baseadas em tokens JWT (JSON Web Tokens). Essa abordagem visava garantir uma autenticação e autorização robustas e stateless entre os backends e os frontends, mitigando vulnerabilidades no consumo das informações indexadas.

Gestão de Repositórios e Configurações: Uma ação metodológica fundamental foi a reestruturação do código-fonte. Realizou-se a separação das aplicações, migrando-as do repositório monolítico "data-ingestion" para um repositório exclusivo denominado "infra". Nesse novo repositório, foram aplicadas as respectivas permissões de acesso e configurados de forma precisa os volumes do Docker, garantindo a persistência correta dos dados processados pelas camadas Bronze, Prata e Ouro do Datalake.

3 RESULTADOS ALCANÇADOS

As atividades desenvolvidas culminaram na estabilização e validação da arquitetura tecnológica, permitindo o cumprimento da Meta 6. Os principais resultados alcançados incluem:

- **Ambiente de Produção Operacional**: A infraestrutura baseada em Docker Swarm e Portainer encontra-se estável, suportando adequadamente as rotinas de ingestão (Logstash), agendamento (Airflow), processamento (Spark) e indexação (Elasticsearch) estipuladas na arquitetura do projeto.
- **Segurança Aprimorada**: A proposta de adoção de tokens JWT estabelece um canal de comunicação seguro entre os backends (APIs REST) e as interfaces de consumo (Painel de Estatísticas e Portal de Serviços), garantindo que apenas requisições autenticadas e autorizadas acessem os dados sensíveis do INPI.

- **Organização e Manutenibilidade do Código**: A criação do repositório específico de infra, dissociado do data-ingestion, resultou em uma governança de código muito superior. As definições de volumes, redes e permissões estão agora devidamente versionadas e isoladas, facilitando futuras manutenções, implantações contínuas e reduzindo o risco de conflitos entre a equipe de infraestrutura e a equipe de engenharia de dados.

4 CONSIDERAÇÕES FINAIS

O período de vigência da bolsa foi essencial para transformar uma arquitetura conceitualmente definida em um ambiente de produção resiliente, seguro e bem governado. As discussões e implementações em torno das tecnologias de orquestração (Docker Swarm/Portainer), processamento de dados (Airflow/Spark/Elasticsearch) e segurança (JWT) provaram-se acertadas para os desafios do projeto Serviços PI - INPI.

A segregação dos repositórios consolida uma maturidade operacional importante, deixando um legado de organização que beneficiará as próximas fases do projeto. Dessa forma, as atividades previstas no plano de trabalho, em especial a implementação do novo sistema em ambiente de produção (Meta 6), foram plenamente atingidas.

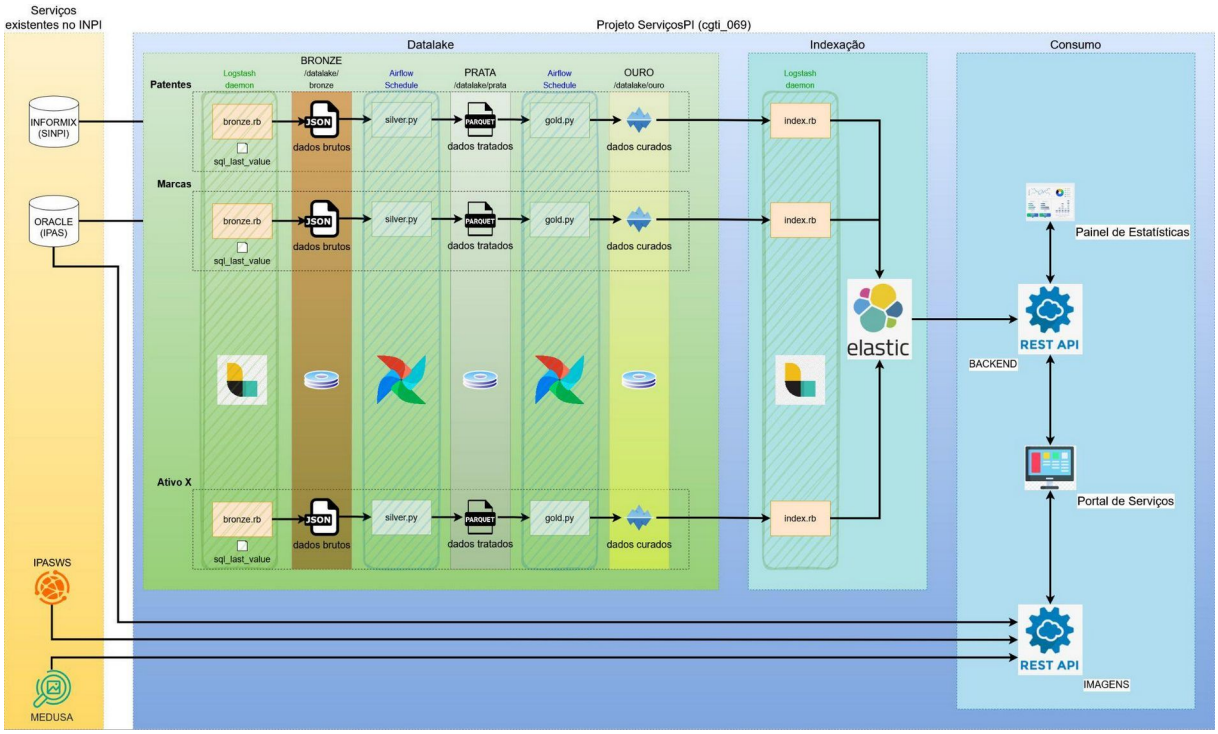
REFERÊNCIAS

DATABRICKS. What is the medallion lakehouse architecture? Databricks. Disponível em: https://docs.databricks.com/en/lakehouse/medallion.html. Acesso em: 13 set. 2026.

ARMBRUST, M. et al. Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics. In: CIDR 2021 - 11th Conference on Innovative Data Systems Research, 2021.

Anexos - documento ou material não elaborado pelo pesquisador, utilizado para comprovação, fundamentação ou complementação.

ANEXO A – Arquitetura Medalhão proposta no Projeto



De acordo:

Documento assinado digitalmente
RODRIGO GONCALVES DE BRANCO
Data: 30/09/2026 11:46:04-0300
Verifique em https://validar.iti.gov.br

Rodrigo Gonçalves de Branco
Pesquisador

Documento assinado digitalmente
WASHINGTON LUIS RIBEIRO DE CARVALHO SEG
Data: 08/10/2026 06:03:23-0300
Verifique em https://validar.iti.gov.br

Washington Luís Ribeiro de Carvalho Segundo
Coordenador do projeto