

Postmortem da Configuração do Apache Airflow

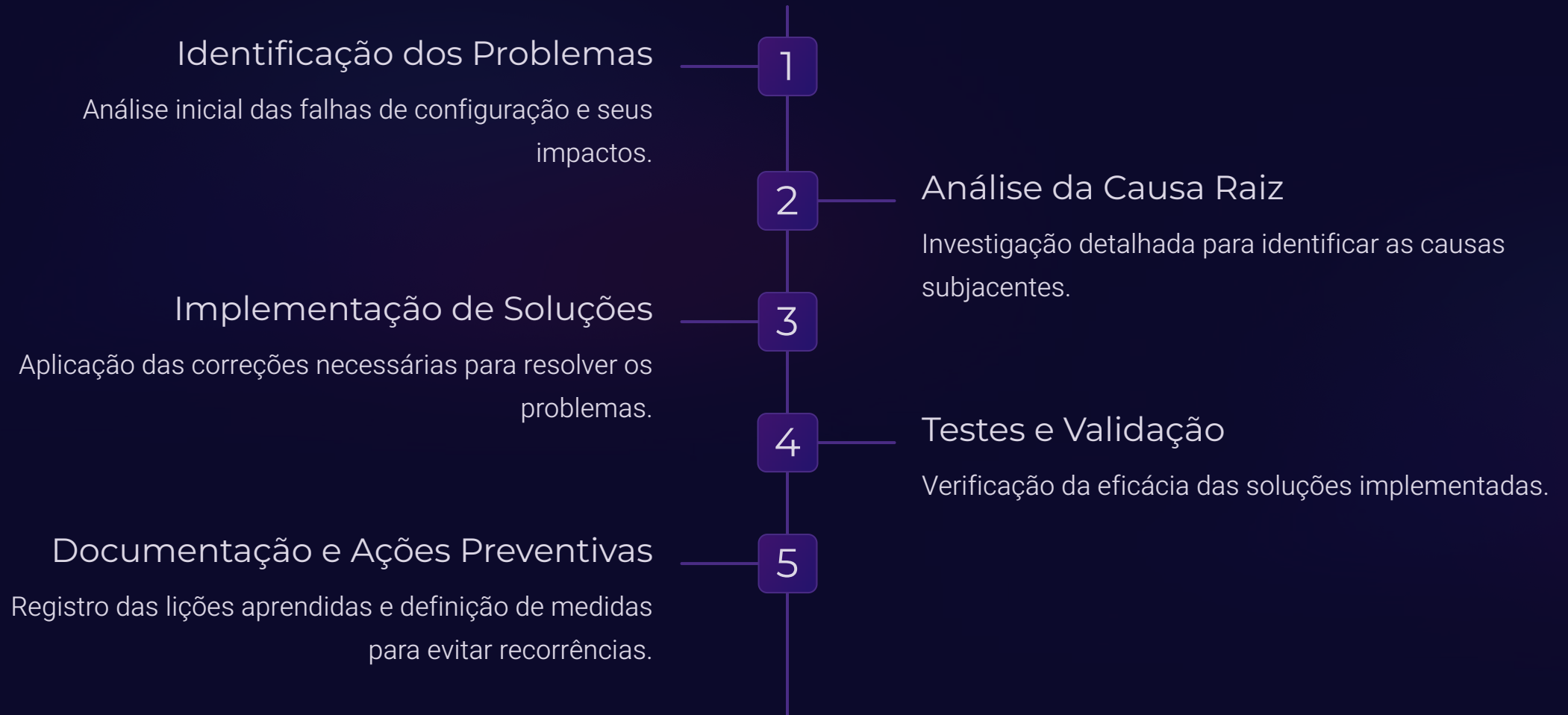
Este documento detalha os desafios e soluções encontrados durante a configuração inicial do Apache Airflow. Nosso objetivo é fornecer um registro claro dos problemas enfrentados, suas causas raízes e as ações corretivas implementadas, garantindo uma infraestrutura Airflow robusta e estável para o futuro.



by **Pedro Amarante**



Cronograma do Postmortem da Configuração do Airflow



Causas Raiz e Soluções Detalhadas

Problemas de Autenticação no Banco de Dados

A configuração inicial falhou devido a erros de autenticação. O Airflow não conseguia se conectar ao Postgres porque utilizava um usuário incorreto.

- **Erro:** Falha na autenticação da senha para o usuário "airflow".
- **Solução:** Criamos um usuário e banco de dados dedicados ao Airflow no Postgres e ajustamos as configurações de conexão (.env) para refletir as credenciais corretas.

Mapeamento Incorreto de Diretórios

O Airflow dependia de diretórios locais para logs e plugins que não estavam corretamente mapeados.

- **Erro:** Diretórios de logs e plugins ausentes.
- **Solução:** Criamos os diretórios necessários no host e ajustamos o docker-compose.yml para garantir a montagem correta dos volumes, permitindo que o Airflow acesse e utilize esses diretórios.

Healthcheck

O Docker verificava a saúde do container pela porta do healthcheck, como estava errada o container entrava em looping.

- **Solução:** Corrigimos a configuração da porta de healthcheck do container airflow-logs para refletir o valor correto e garantir que o monitoramento da saúde do container seja adequado.
- **Erro:** A porta de healthcheck do container airflow-logs estava configurada incorretamente.

Usuário do container

O usuário que estava sendo utilizado no container não era o correto.

- **Erro:** Usuario errado causando problemas de permissionamento dentro do airflow.
- **Solução:** Ajustamos o usuario dos containers e as permissões para ele. Assim o airflow não terá mais problemas de permissionamento.

Ações Preventivas e Melhorias Contínuas



Configuração de Usuário e Permissões

Definir corretamente **AIRFLOW_UID** e **AIRFLOW_GID** antes da inicialização dos containers.



Gerenciamento de Diretórios

Criar e configurar os diretórios **dags**, **logs** e **plugins** com as permissões adequadas para evitar problemas de acesso.



Testes Automatizados

Implementar uma pipeline de testes para validar a sintaxe das DAGs antes de sua implementação.



Observabilidade e Monitoramento

Monitorar continuamente a saúde da stack e do postgres.



Docker image

Não utilizar imagens 'Latest', as atualizações automaticas podem quebrar o ambiente.



Escalabilidade

Implementar a stack do airflow em um ambiente com recursos suficientes para que as dags possam escalar sem problemas.