



# Controle de Versão de Dados em Projetos de Machine Learning

Roberto Rodrigues Junior

Engenheiro de dados - Jusbrasil  
robertogyn19@gmail.com

# Agenda

- Como gerenciar os dados em MLOps
- Ferramentas existentes
- Exemplos práticos com DVC
- Dúvidas



# Problemas com dados

- Corrompidos
- Faltando
- Modificados
- Sensíveis



# Como resolver os problemas com dados

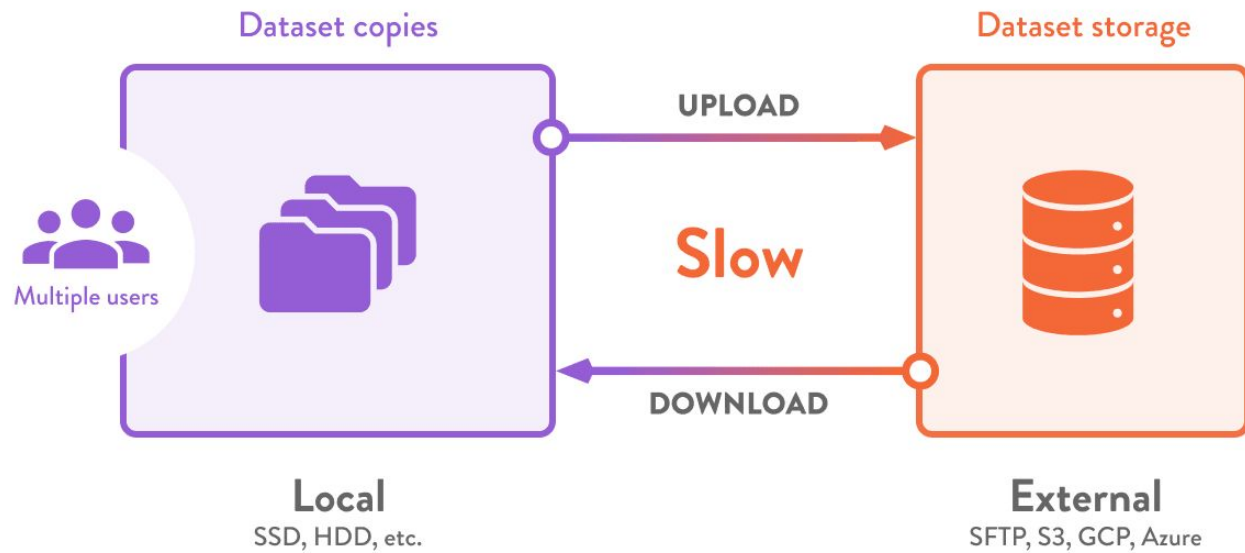
- Processos automatizados
- Validação de qualidade
- Versionamento
- Gestão de acesso





# Ferramentas existentes

- Git LFS
- lakeFS
- dolt
- delta
- DVC
- Pachyderm
- Marquez
- Feast



**PROJECT A**



**DEPLOYMENT**  
(Flask, Docker, etc)



**PROJECT B**



**DATA REGISTRY**  
(GIT+DVC REPOSITORY  
GitHub, GitLab, etc)



**DATA STORAGE**  
(SSH, S3, GCS, etc)



Dúvidas?