

# My-Feast - Implementando uma Feature Store em um ecossistema de dados híbrido (On-Prem & Cloud)

## Overview

Este projeto é uma demonstração de uso de uma feature store em uma arquitetura de dados híbrida, composta por um data lake on-premises (Hadoop/HDFS) e um data lake na cloud (Azure Data Lake Gen2).

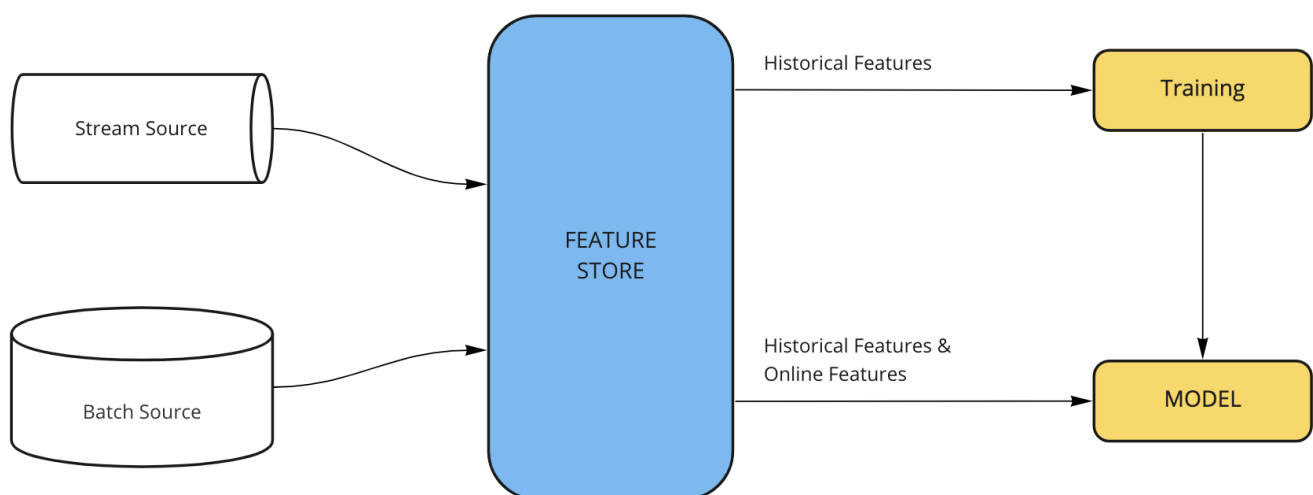
## Conceitos e Objetivos

'Deploying ML is hard.'

A máxima acima, talvez exagerada (ou nem tanto), é o principal motivo da existência das Feature Stores. Em resumo, uma **feature store** é um sistema responsável por gerir, "armazenar" e servir dados, pré-processados, para modelos de ML.

Em um pipeline, ou ciclo de vida, de um modelo, os processos de Feature Engineering e produção das features criadas demandam tempo e possuem grande complexidade; um cientista de dados atuando na criação de um conjunto de entidades e suas variáveis, não é capaz de, rapidamente e com autonomia, publicar suas features para treino e serving destes dados aos modelos de ML.

Uma feature store soluciona este problema, ao atuar como uma camada única de acesso aos dados, desacoplando o storage das features da recuperação destas. Em essência, uma feature store complementa o ecossistema de dados existente, ao fornecer uma camada de **gestão de metadados** das features que permite a **Descoberta**, **Compartilhamento** e **Reuso** de fetures em uma organização, além de servir dados de maneira **Consistente no tempo (point in time correctness data)**, evitando a ocorrência de desvios entre os dados de treino e serving do modelo (**Training-Serving skew**) e facilitando o processo de **Deploy de novas features**.



Neste projeto foi criada uma arquitetura híbrida, contando com um cluster **Spark Standalone** e um **HDFS**, ambos dockerizados e um lake na cloud, utilizando como Storage o **Azure Data Lake Gen2** e como engine

de processamento e consumo analítico, a plataforma [Azure Databricks](#). A feature store utilizada neste projeto é o [Feast](#), um projeto Open Source, agnóstico em relação à infraestrutura.