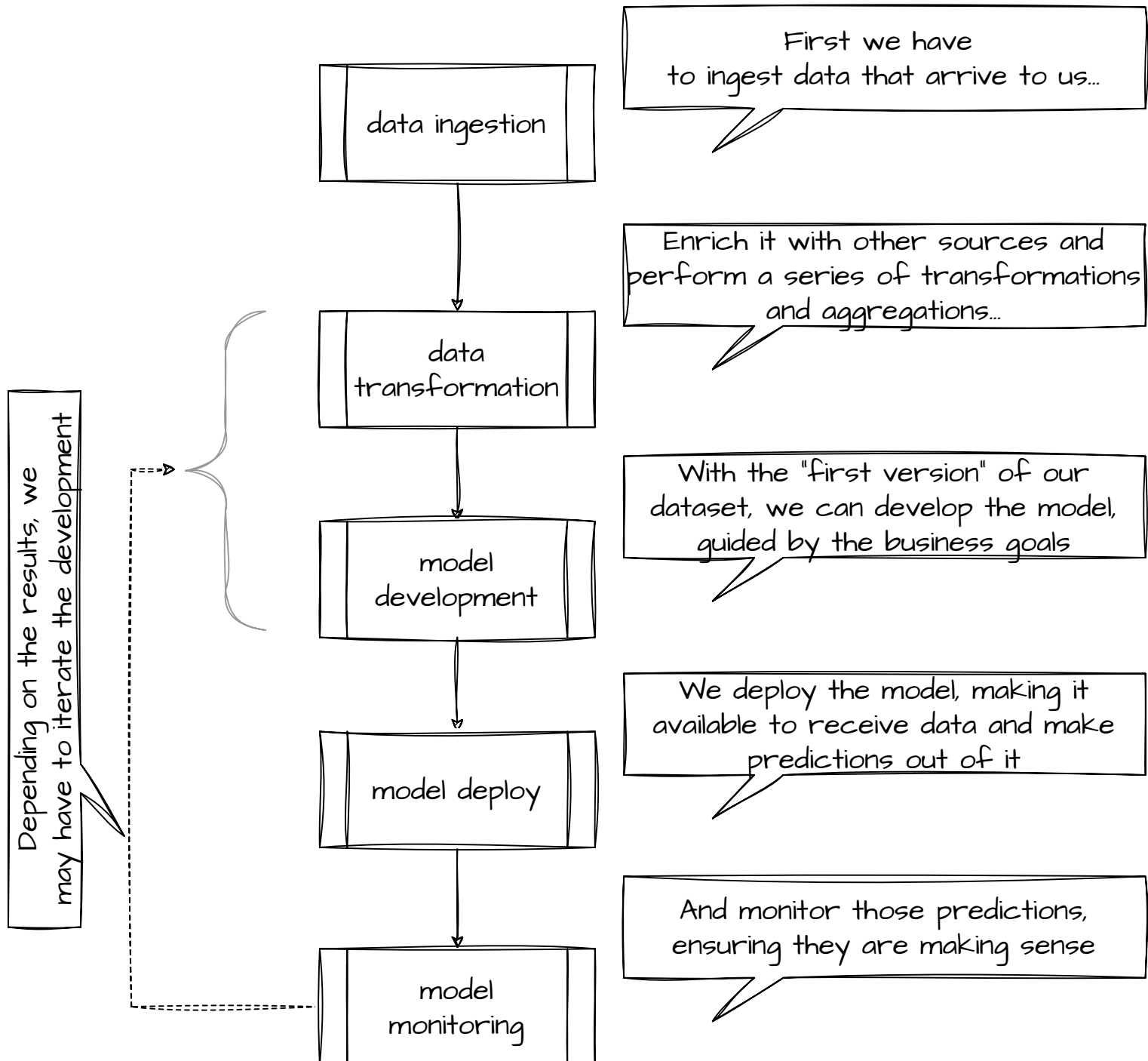


General machine learning pipeline

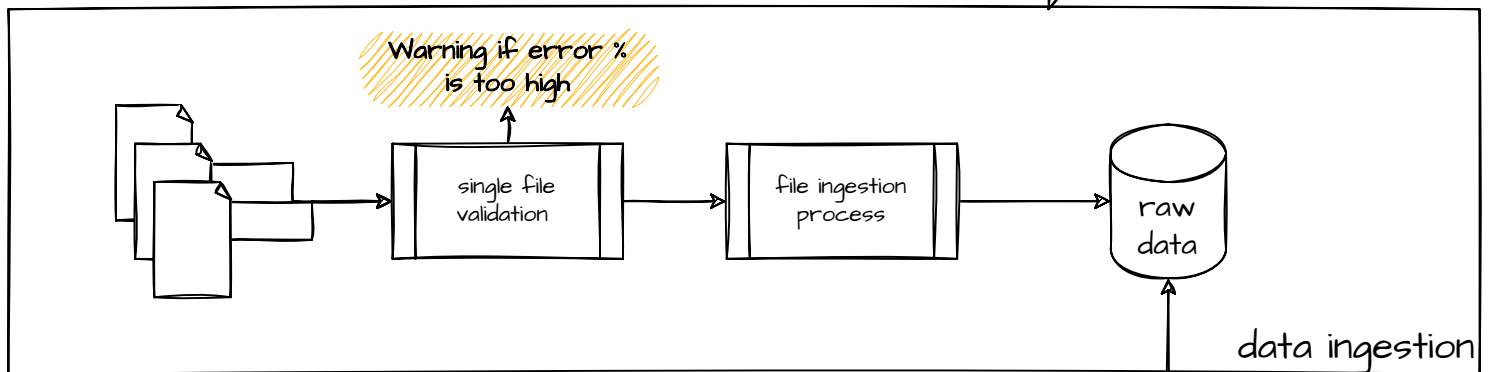
Here we give a *bird's eye view* of the proposed machine learning platform. In the next sections, a more detailed exploration of how it works will be given.



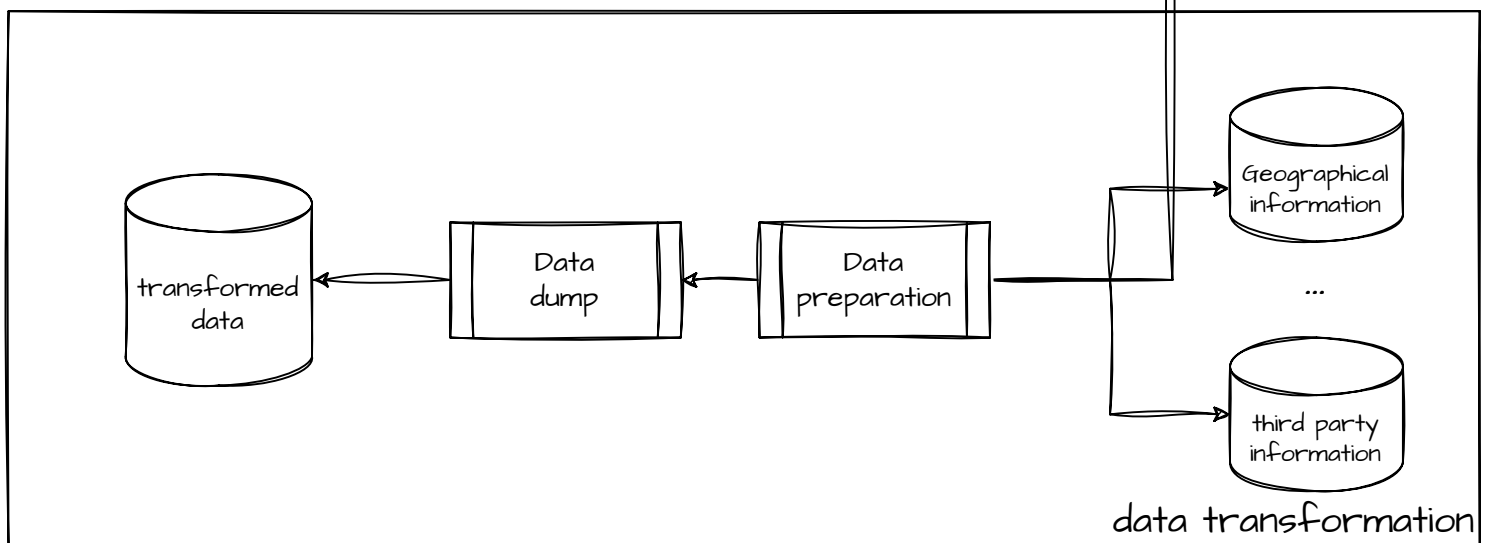
Feeding data to our model

First we have to have clean data to consume!

A series of files containing the data arrive and we perform some validations, to ensure they are correctly formatted and contain the expected fields. After that, we perform a process to ingest data into a raw database.



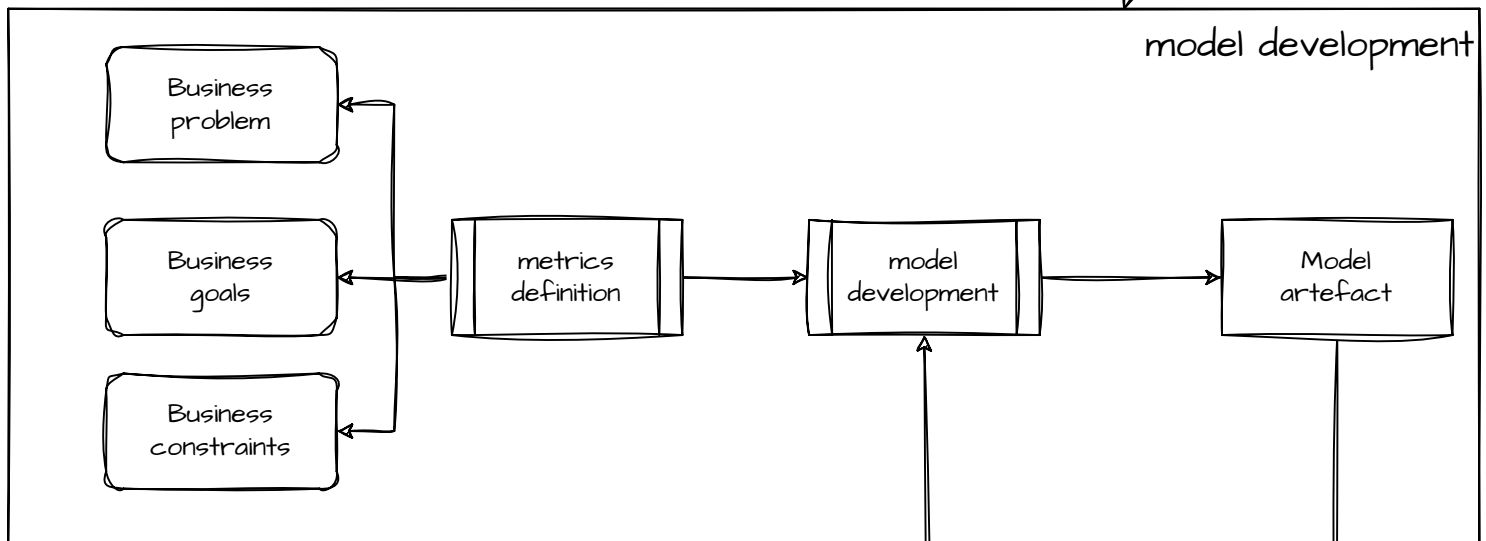
Once we have ingested the data, we can merge fields from other datasets of interest, clean data, run all steps that involve aggregations, statistical calculations, and more. In the end, we save the transformed dataset to be consumed by the model.



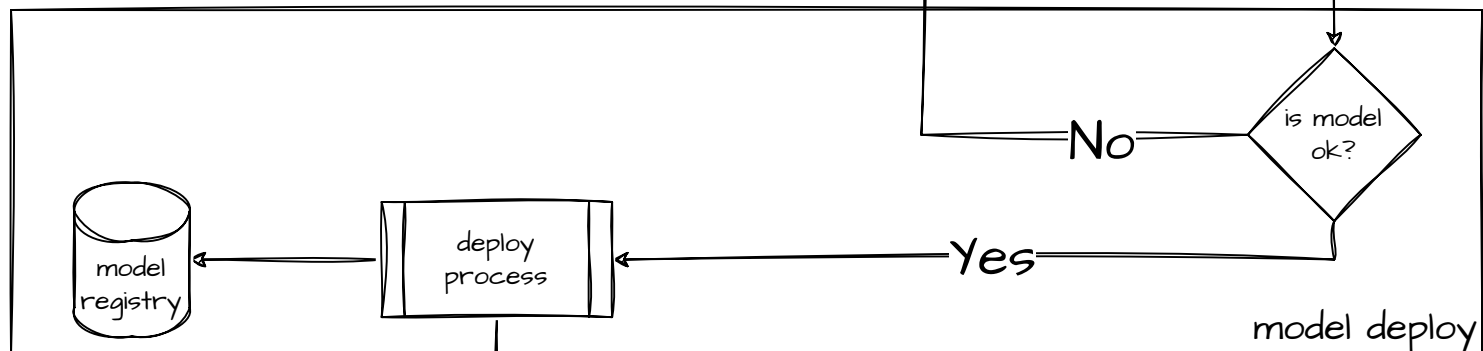
Model development

It's time to develop our model!

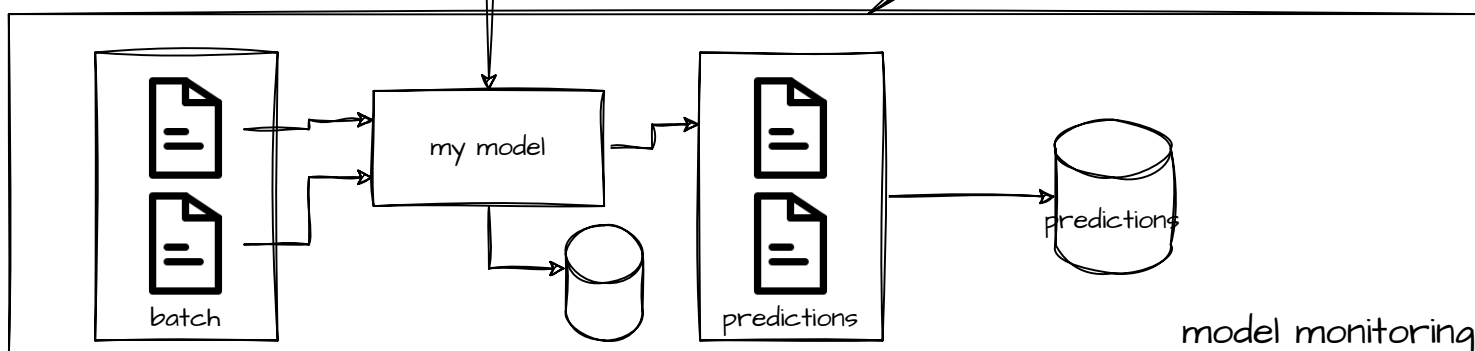
Guided by the business problem and goals, we start modeling. We test different algorithms, features, and optimizations... We may find that we have to go back to the previous steps...



If the model passes the quality tests, we can run the deploy process and save the details of the deployed model on a model registry

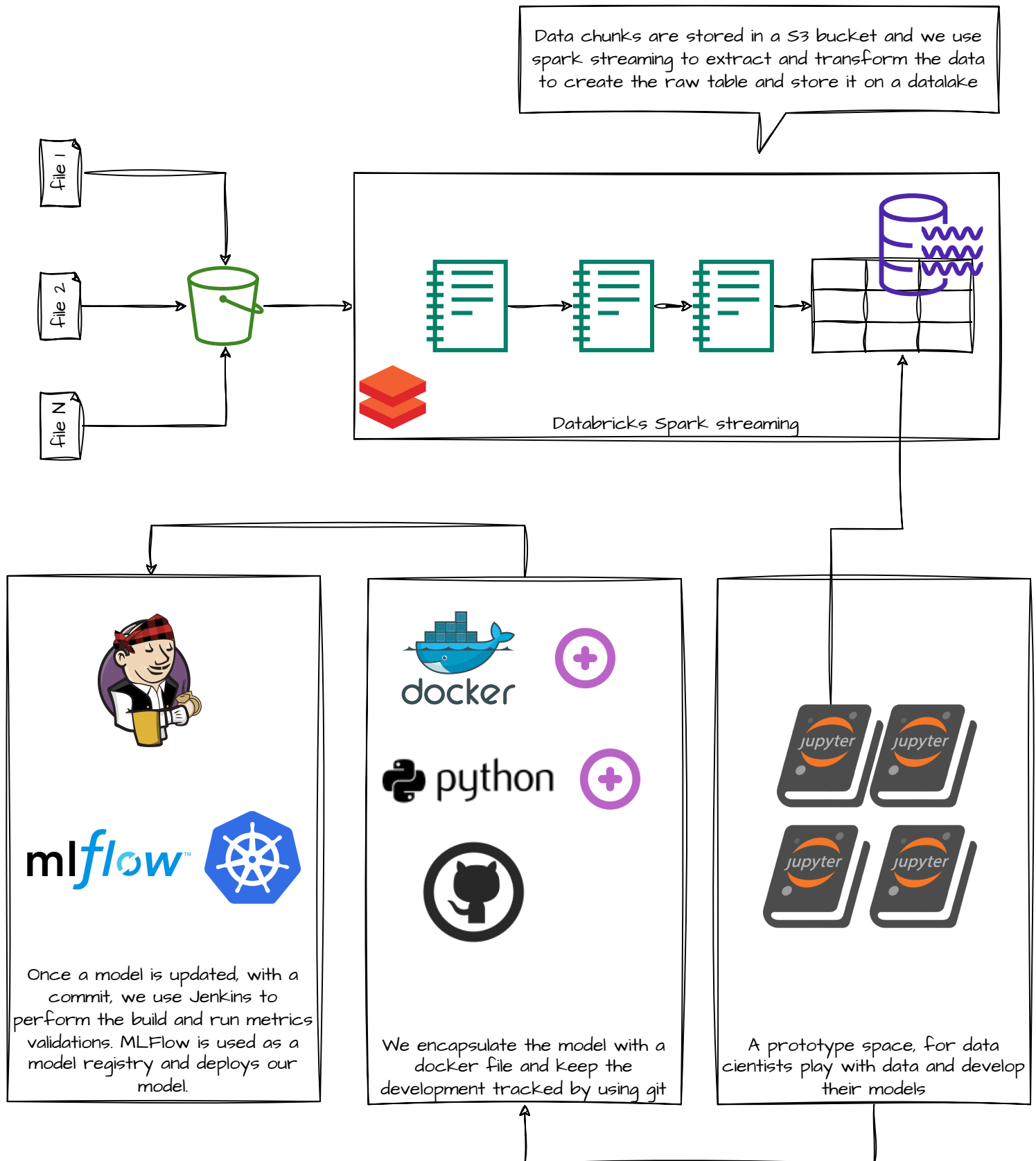


We use the model to make predictions of the data batch and store the results to analyse model quality, perform audits...



Tech stack for deploying a model

We develop here our tech stack (we try to keep only open-source projects).



Tech stack for a real-time application

How to allow real-time predictions

