Test mini dataguru

Build a simple web API to help a Machine Learning team organize its data. The data consist of image that have tags, those tags are the labeling of the image. For exemple an image can be labeled with "Dog" if there is a dog in the image.

IMPORTANT:

To simplify the job, you can fake all the data: for exemple there are images involved but no need to really handle image upload, you can just fake the image name, size, url etc...

Expected work

A github repo with the code for a web API, REST if possible (GraphQL also accepted). You can deploy it if you want (it's neat, but do not spend money on this!) but in any case you should provide clear steps to be able run it locally, so we can look at the code AND test it with curl (or a postman).

It's better if it is in Python (Flask). If you are not familiar with our stack, you can go with a standard Node.js framework.

For the DB we recommend a simple SQL db. SQLite or PostgresQL.

It is better if your project is containerized.

User Stories:

1. AADS (As a Data Scientist) I can upload an image, it stores a url, a name, a type.

2. AADS I can paginate on all the image I have uploaded

3. AADS I can create new Tags, the tags only have a name. (For exemple a tag "Cat")

4. AADS I can list all of the existing Tags

5. AADS I can add a tag to an image to signify that this image has the tag "Cat"

6. AADS I can list all the tags of an image.

7. AADS I can paginate all the images that have a tag.

Bonus (in order of priority)

1. AADS I can group image in datasets (a collection of images). Image can be in one or multiple dataset and a dataset can have one or multiple image. I can create new dataset and add and remove images from datasets.

2. [HARD] AADS I can have different "versions" of labeling for an image. For exemple for a given image, it can have a first labeling with no tags (the person did not see anything), another version with the tag "Dog" and "Garden" (the person saw a dog and the image was in a garden) and another version with the tag "Wolf" and "Forest" (the person thought it was actually a wolf and in a forest). In that case you cannot just add tags one by one to the image as before, but you add the whole "version" in one go with all the tags in the version. For exemple I have my image, it has a version with tags "Tree" and "Butterfly", I also see a cat on the image, I add a new version that has tags "Tree", "Butterfly" and "Cat" in one operation.

3. AADS my API has a swagger

Open question (Bonus/Hard)

In Machine Learning, the model need to knows the tags in an image (This image has a Cat and a Dog), but also needs to know what was NOT in the image, despite the fact that the model can detect it. For exemple if we want to create an AI that can detect Cat, Dogs and Birds, we need to know that the model is looking for "Cat" "Dogs" and "Birds" and that if the image only has a "Cat" and "Dog" tag, it means there a no birds on the picture. If we have a model that only detect "Cat" and "Dog", and the image has a "Cat" and "Dog" tag, it does not mean there a not birds in the picture, there might be one or there might not, it only means there is a Cat and a Dog and we are not interested in knowing if there is a bird.

How would you go about implementing this ? (Do not do it, just explain your ideas)