

From Jupyter Notebook to **reproducible** ML pipeline

Rob de Wit – 2023-04-21
PyCon US, Salt Lake City

Rob de Wit

Utrecht, the Netherlands

Developer Advocate*

Water-Pokémon trainer



Goal

**Generate Pokémon
with Stable Diffusion**

Approach

**Use Stable Diffusion,
train LoRA on Pokémon**

Implementation

Prototype in

Jupyter Notebook





Experimentation

Find the best Pokémon

Experimentation

Find the best predictions

Experimentation

Find the best model

IRL

We experiment all the time

Reproducibility

Track all components

Experiment =
data + code + params



+



DVC

1 Data version control

2 Pipelines

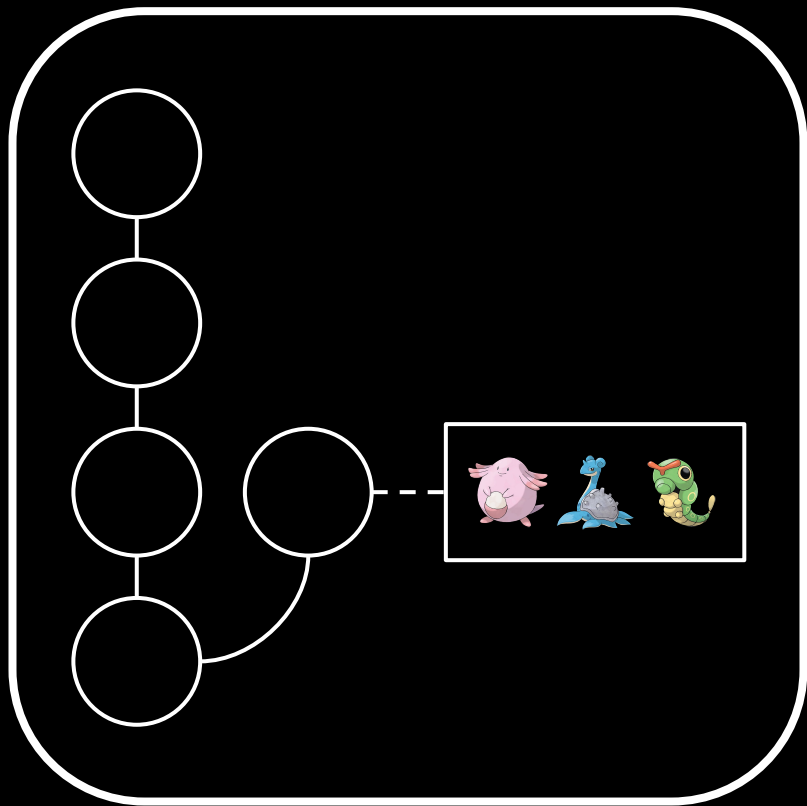
3 Experiments

DVC

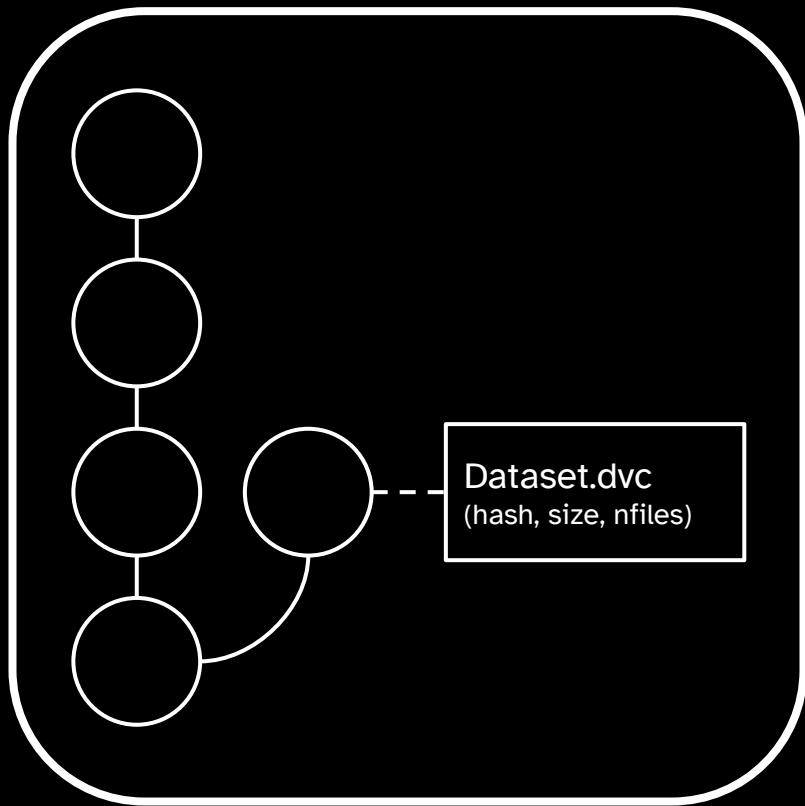
1 Data version control

2 Pipelines

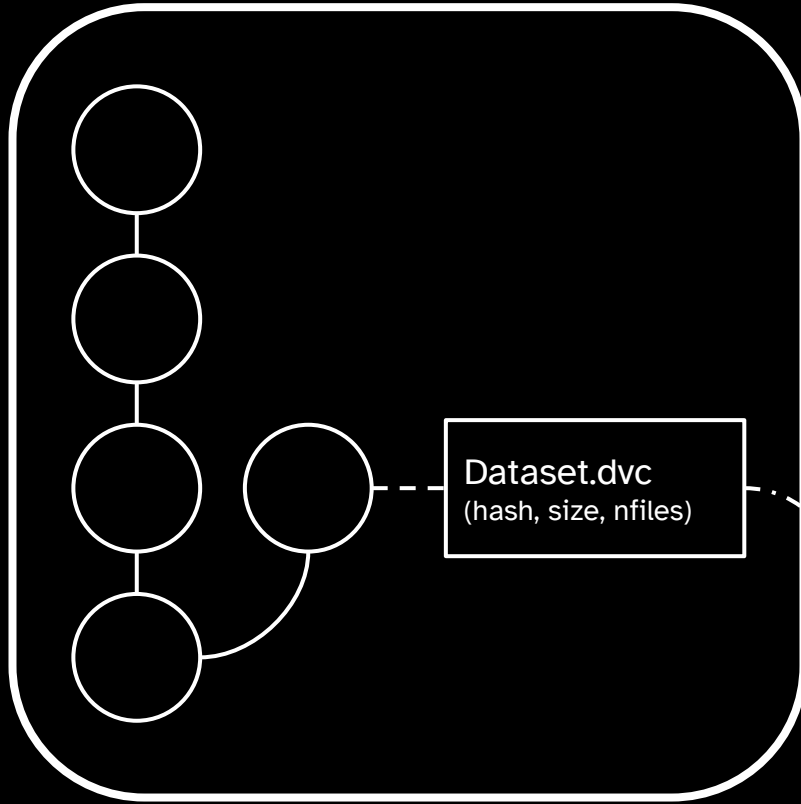
3 Experiments



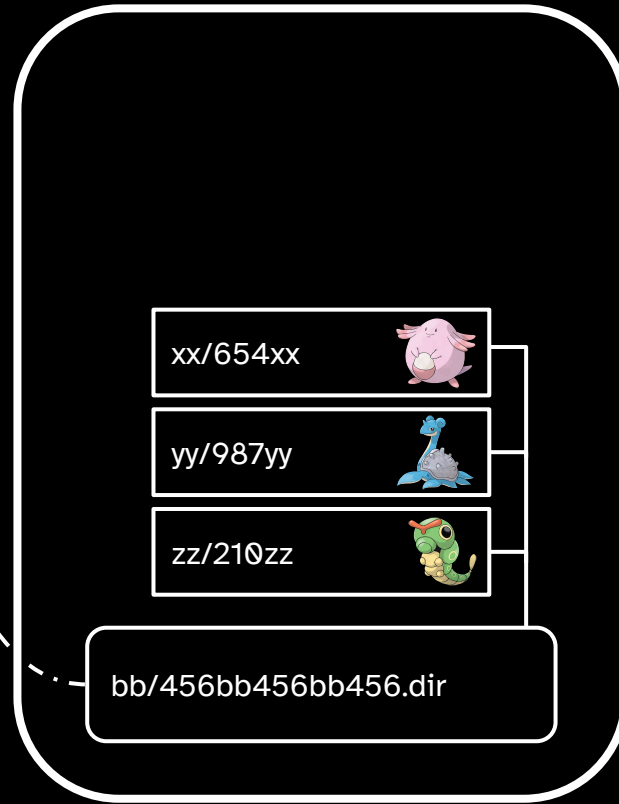
Git repo



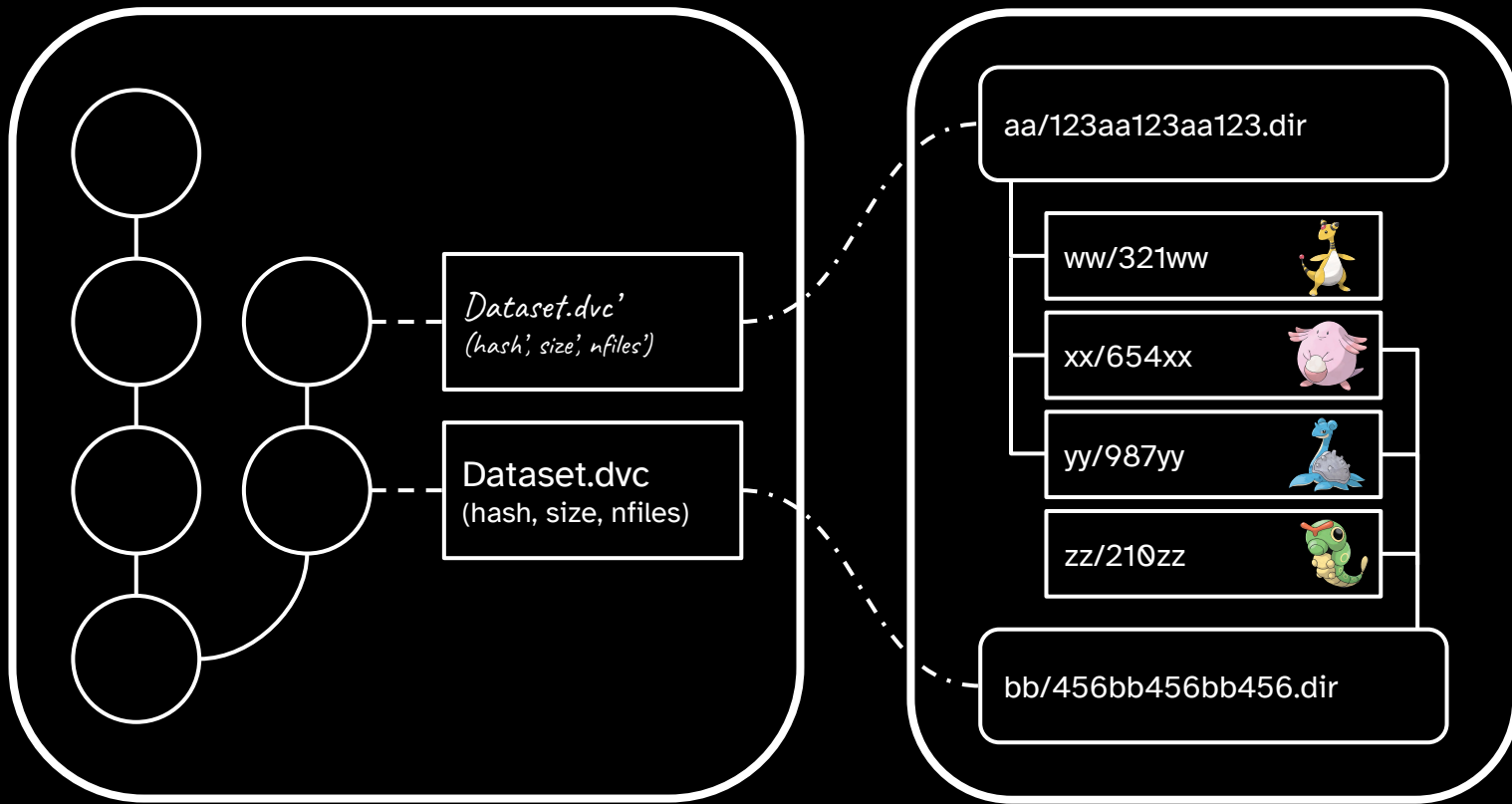
Git repo



Git repo

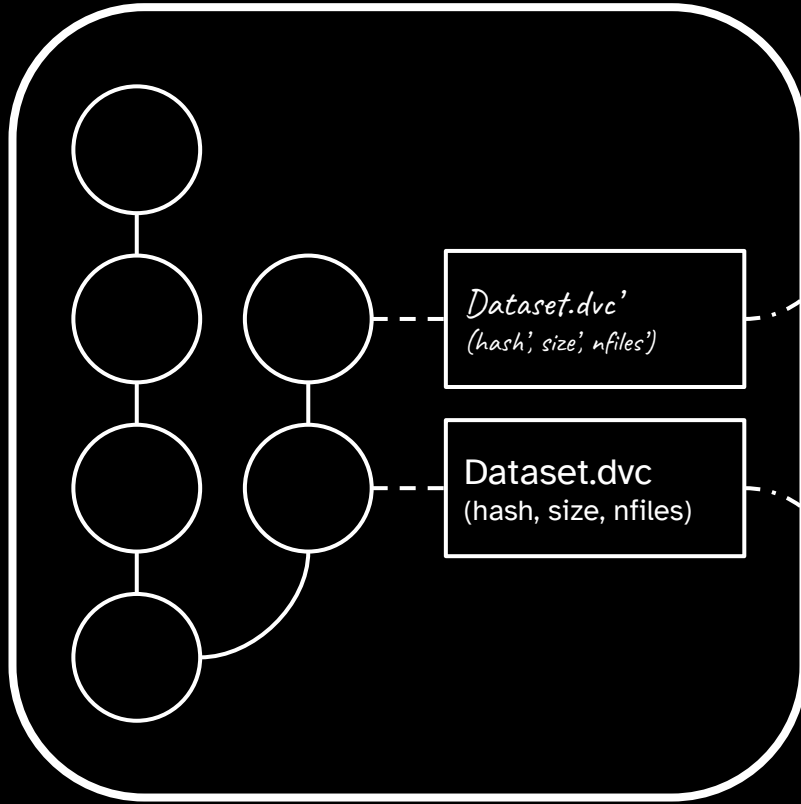


`.dvc/cache`

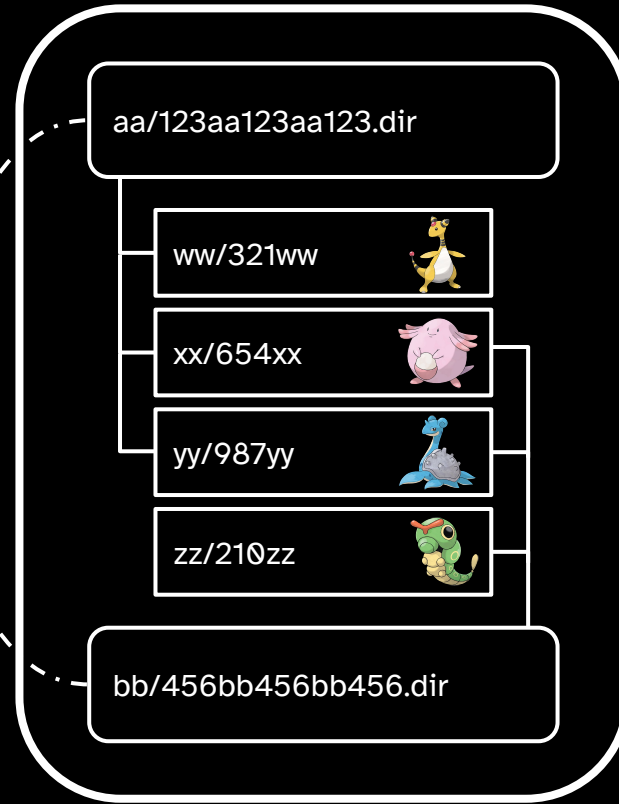


Git repo

.dvc/cache



Git repo



.dvc/cache



Remote
Storage

DVC

1 Data version control

2 Pipelines

3 Experiments

set_up_diffusers

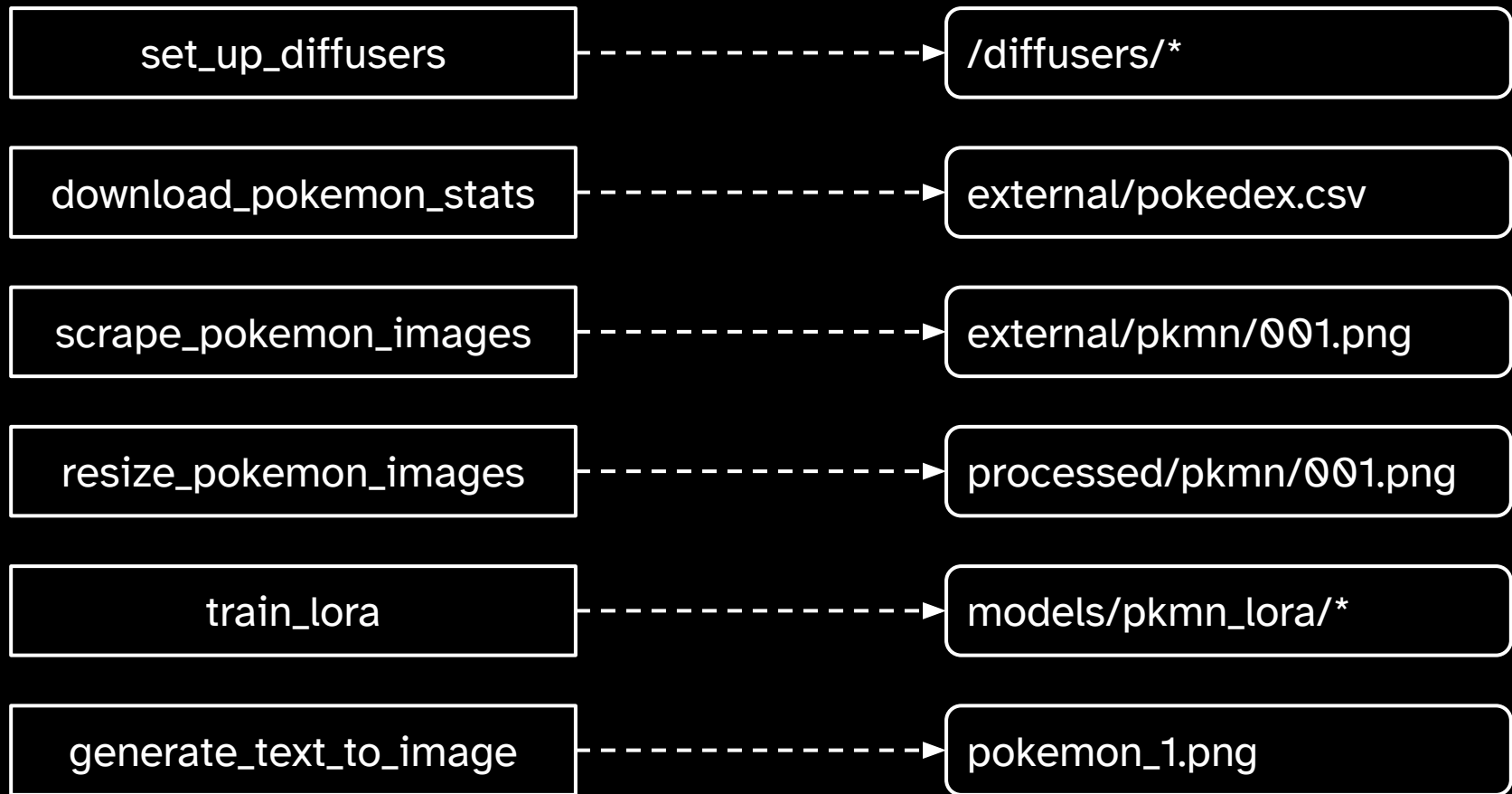
download_pokemon_stats

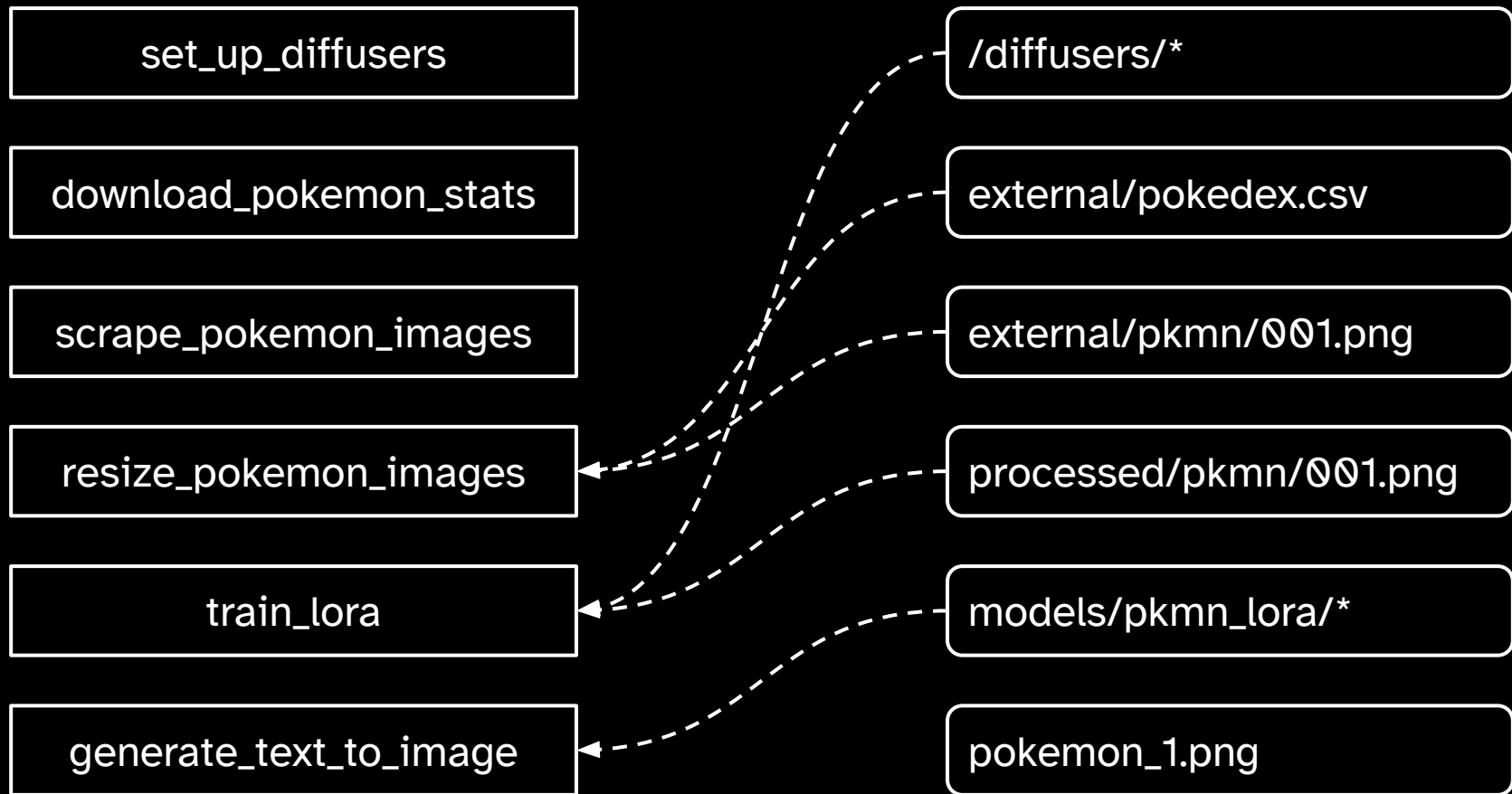
scrape_pokemon_images

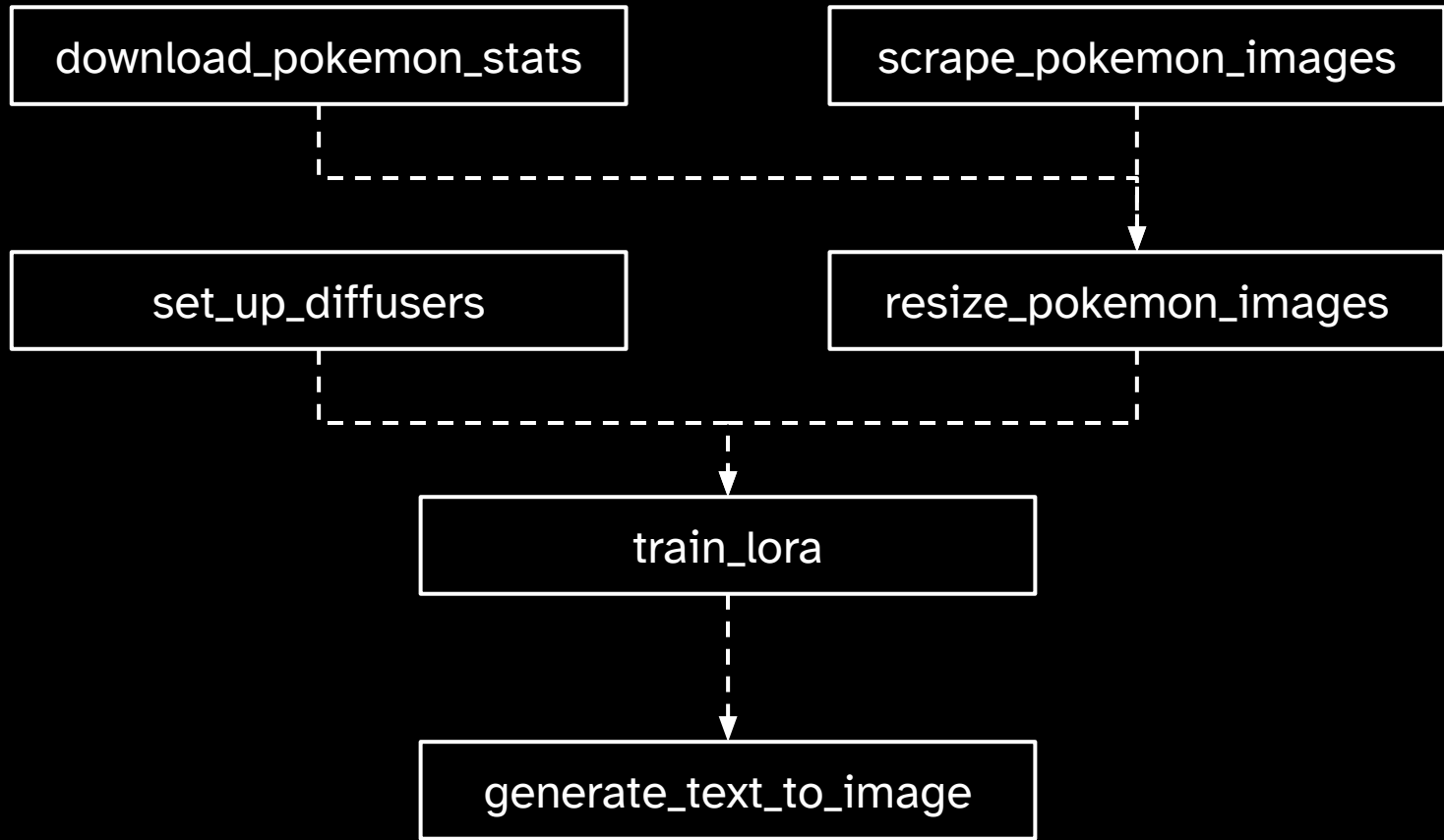
resize_pokemon_images

train_lora

generate_text_to_image





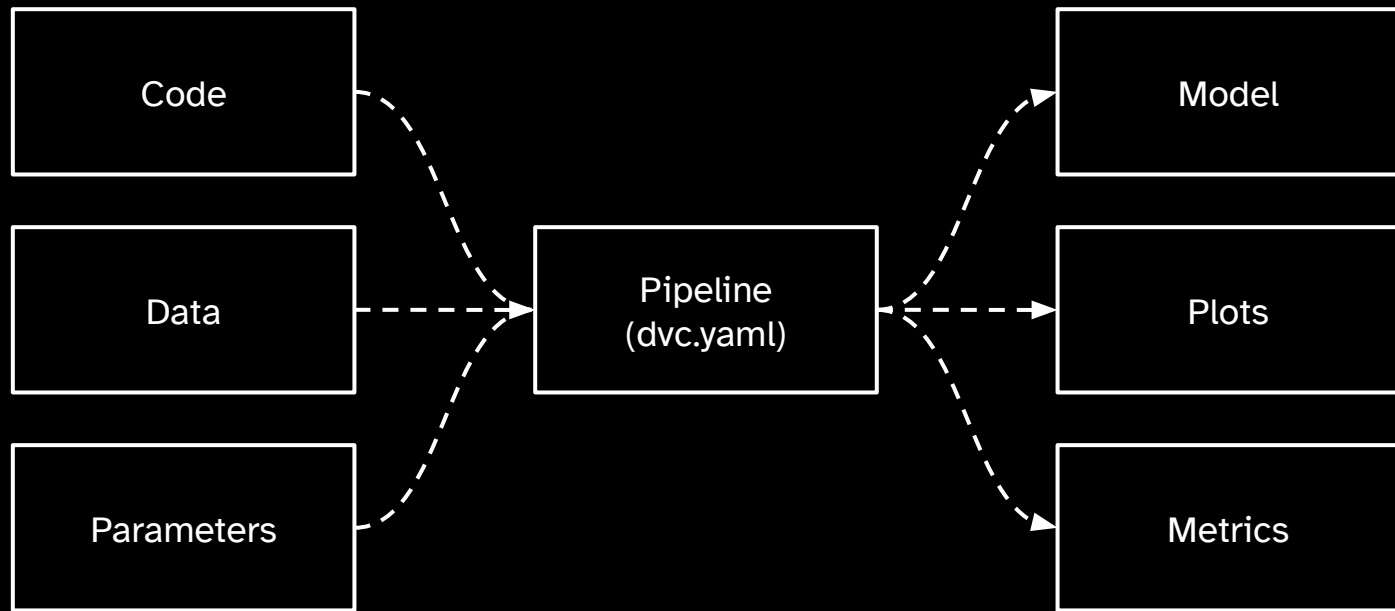


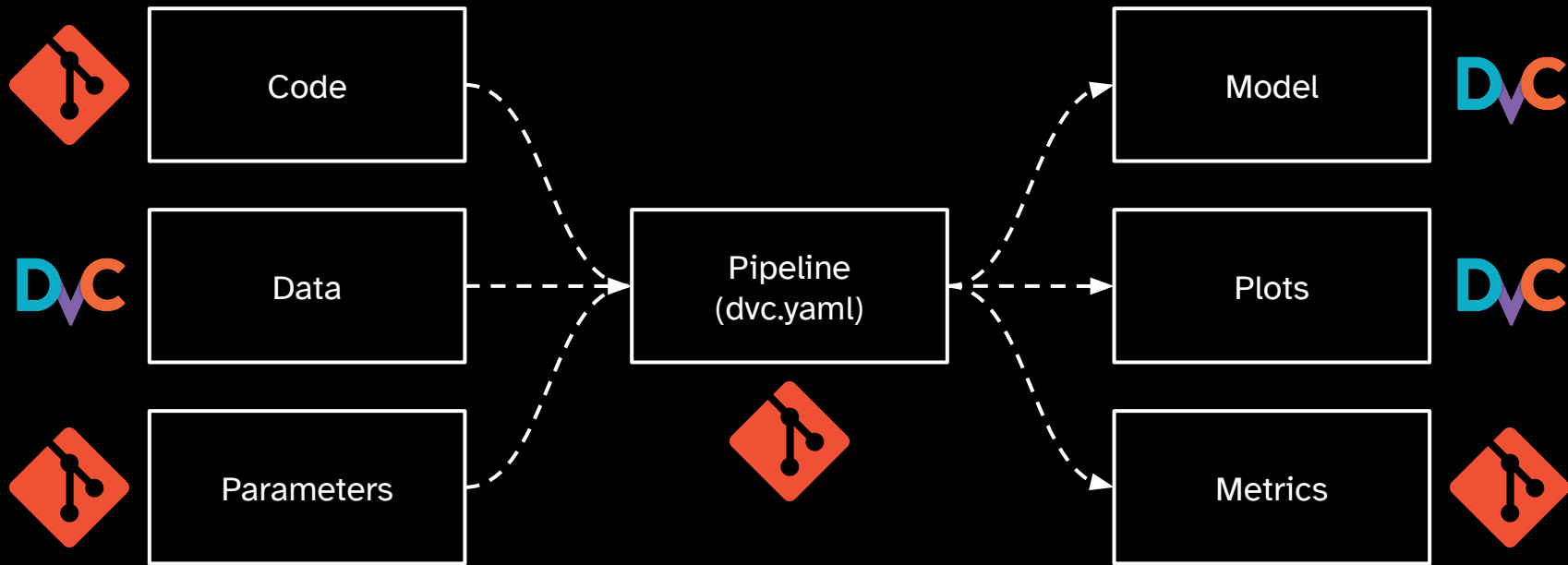
DVC

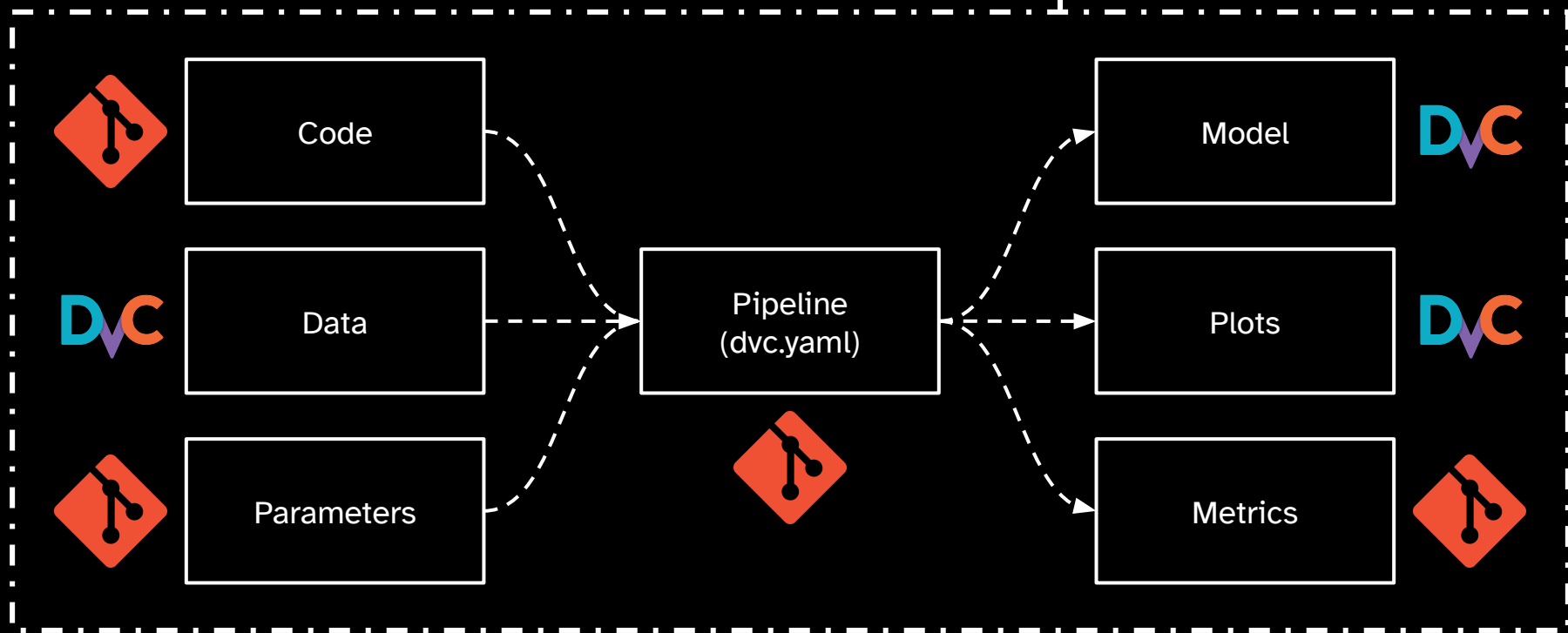
1 Data version control

2 Pipelines

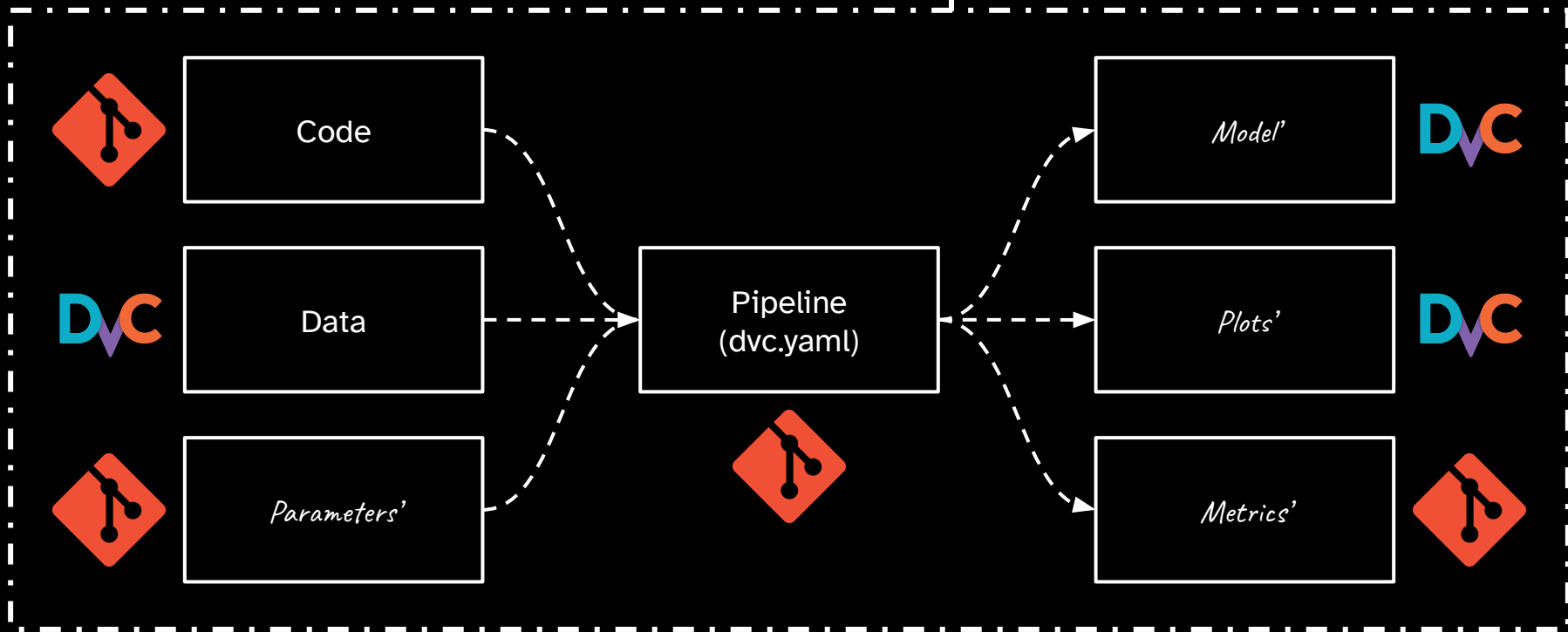
3 Experiments







Experiment (d69eb2f)



Experiment (f27ec9a)

Implementation 2

DVC pipeline

Recap and takeaways

**In data science we
experiment constantly**

**Experimentation requires
reproducibility**

**Jupyter Notebooks are
great for **prototyping**, not
great for reproducibility**

**We can achieve full
reproducibility with
Git and DVC**

**Together, Git and DVC track
data, code, and parameters**

We can use DVC for
Data version control

We can use DVC for
Pipelines

We can use DVC for
Experiment management

Thank you!

robdewit.nl
rob@binary3.dev

#opentowork



Further resources

- dvc.org
- dvc.org/chat
- github.com/RCdeWit/sd-pokemon-generator

Sources used

- dvc.org/doc/
- huggingface.co/docs/diffusers/
- github.com/cloneofsimon/lora/
- replicate.com/blog/lora-faster-fine-tuning-of-stable-diffusion/
- aituts.com/stable-diffusion-lora/
- huggingface.co/blog/lora0/
- old.reddit.com/r/StableDiffusion/comments/1171zhk/how_can_i_make_a_lora_model_on_my_m1_mac/jerageb/
- civitai.com/models/5115/pokemon-lora-ken-sugimori-style/
- kaggle.com/datasets/brdata/complete-pokemon-dataset-gen-iiiv/
- <https://stable-diffusion-art.com/lora/>