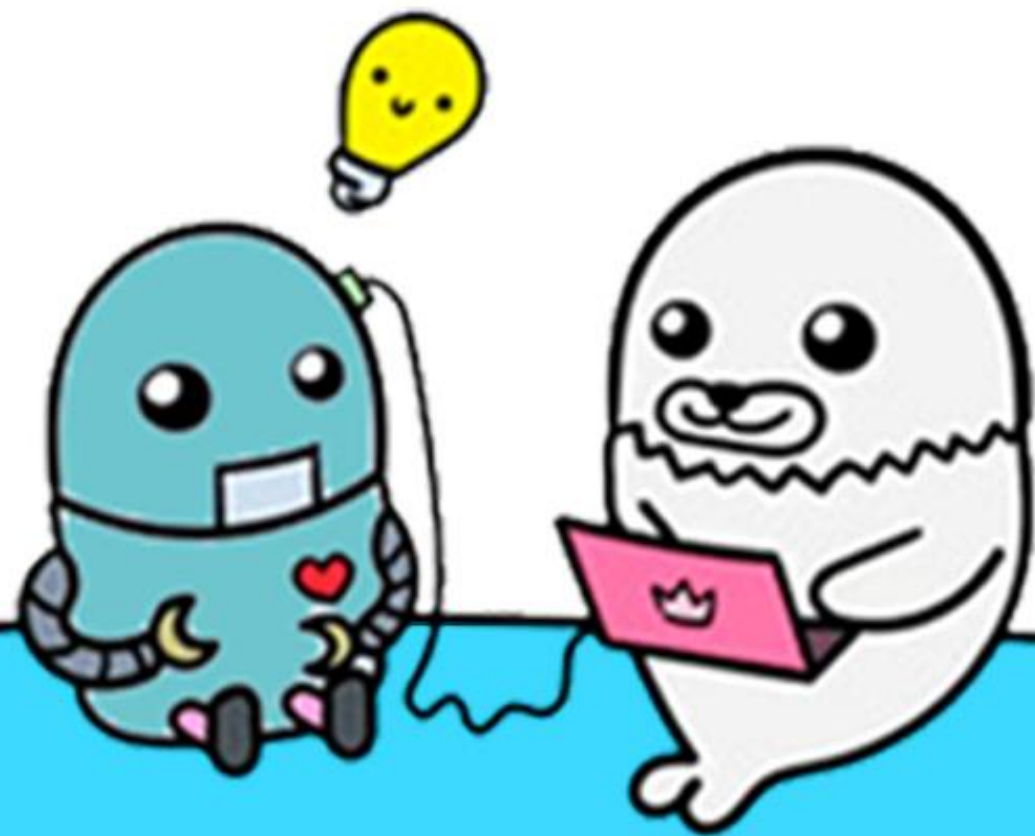


MLOps и production подход к ML исследованиям



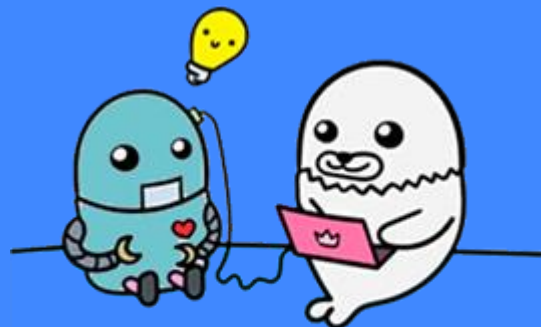
28 марта - 28 мая



Опыт развёртывания ML сервиса в production на примере проекта FindMyBike

Инструменты автоматизации ML исследований, DVC + MLFlow

Антон Ганичев
Газпромнефть ЦР
Руководитель направления NLP
t.me/pavel_kikin





MLflow

MLflow is an open source platform to manage the ML lifecycle, including experimentation, reproducibility, deployment, and a central model registry. MLflow currently offers four components:

MLflow Tracking

Record and query experiments: code, data, config, and results

[Read more](#)

MLflow Projects

Package data science code in a format to reproduce runs on any platform

[Read more](#)

MLflow Models

Deploy machine learning models in diverse serving environments

[Read more](#)

Model Registry

Store, annotate, discover, and manage models in a central repository

[Read more](#)



Mlflow + DVC

DVC

- версионирование данных и артефактов
- воспроизведение экспериментов

Mlflow

- трэкинг экспериментов
- упаковка кода в воспроизводимый формат
- хранение управление и деплой моделей





MLflow Tracking

MLflow Tracking



Listing Price Prediction

Experiment ID: 0

Artifact Location: /Users/matei/mlflow/demo/mlruns/0

Search Runs:

Search

Filter Params:

Filter Metrics:

Clear

4 matching runs

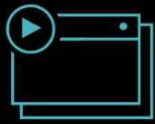
Compare Selected

Download CSV

	Time	User	Source	Version	Parameters		Metrics		
					alpha	lr_ratio	MAE	R2	RMSE
<input type="checkbox"/>	17:37	matei	linear.py	3a1995	0.5	0.2	84.27	0.277	158.1
<input type="checkbox"/>	17:37	matei	linear.py	3a1995	0.2	0.5	84.08	0.264	159.6
<input type="checkbox"/>	17:37	matei	linear.py	3a1995	0.5	0.5	84.12	0.272	158.6
<input type="checkbox"/>	17:37	matei	linear.py	3a1995	0	0	84.49	0.249	161.2



MLflow Projects



Local Execution



Remote Execution



```
my_project/  
├── MLproject  
  
├── conda.yaml  
├── main.py  
├── model.py  
└── ...
```

```
conda_env: conda.yaml  
  
entry_points:  
  main:  
    parameters:  
      training_data: path  
      lambda: {type: float, default: 0.1}  
    command: python main.py {training_data} {lambda}
```

```
$ mlflow run git://<my_project>  
  
mlflow.run("git://<my_project>", ...)
```



MLflow Models

TensorFlow
PYTORCH



APACHE
Spark
dmlc
XGBoost

ML
Frameworks

mlflow
Model Format

Flavor 1

Flavor 2

Standard for ML
models

docker
Inference Code

APACHE
Spark
Batch & Stream
Scoring



Serving Tools

Example MLflow Model

```
my_model/  
├── MLmodel
```

```
run_id: 769915006efd4c4bbd662461  
time_created: 2018-06-28T12:34  
flavors:  
  tensorflow:  
    saved_model_dir: estimator  
    signature_def_key: predict  
  python_function:  
    loader_module: mlflow.tensorflow
```

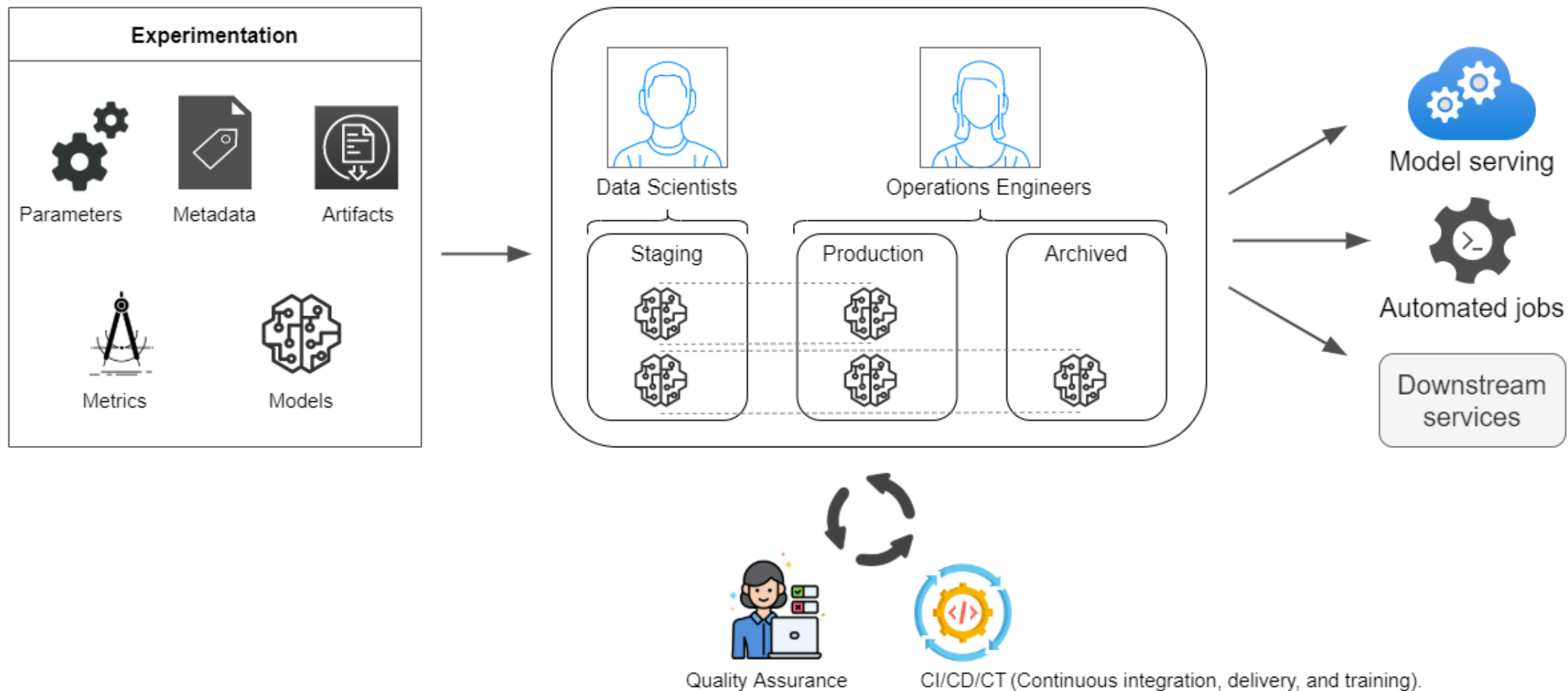
Usable by tools that understand
TensorFlow model format

Usable by any tool that can run
Python (Docker, Spark, etc!)

```
├── estimator/  
    ├── saved_model.pb  
    └── variables/
```

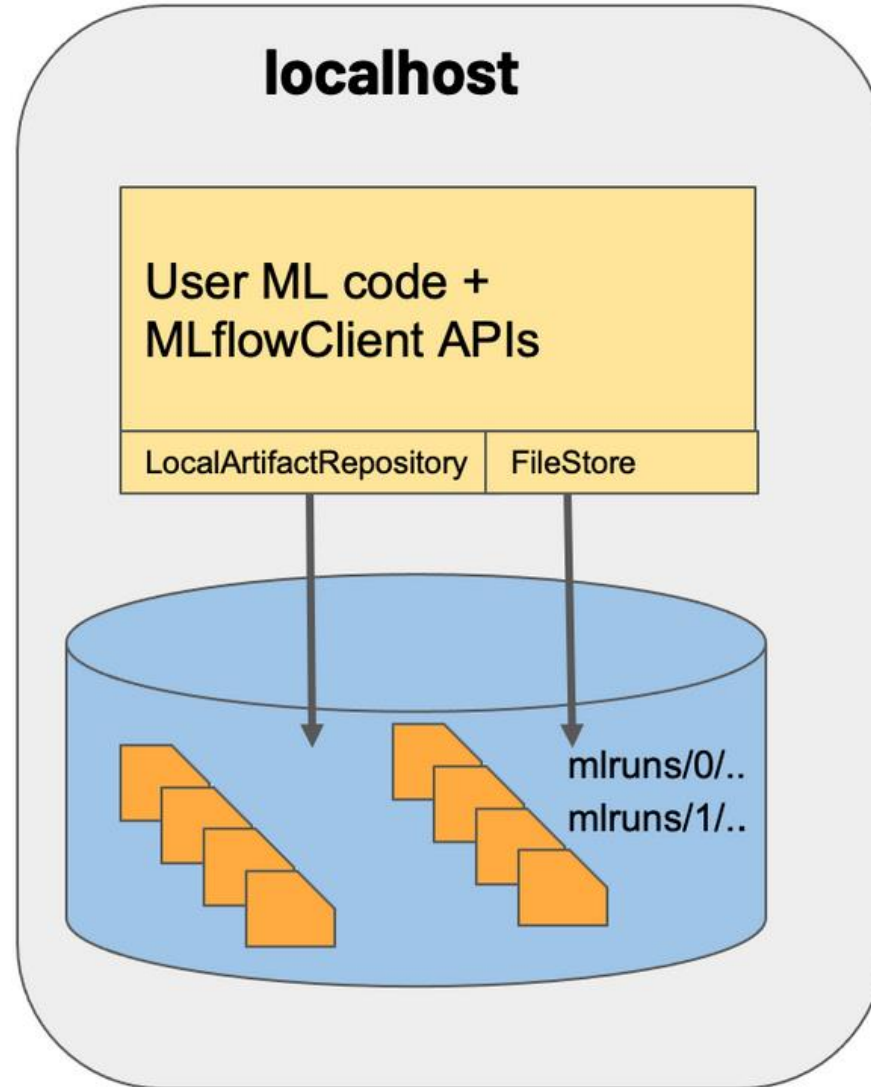



Model Registry



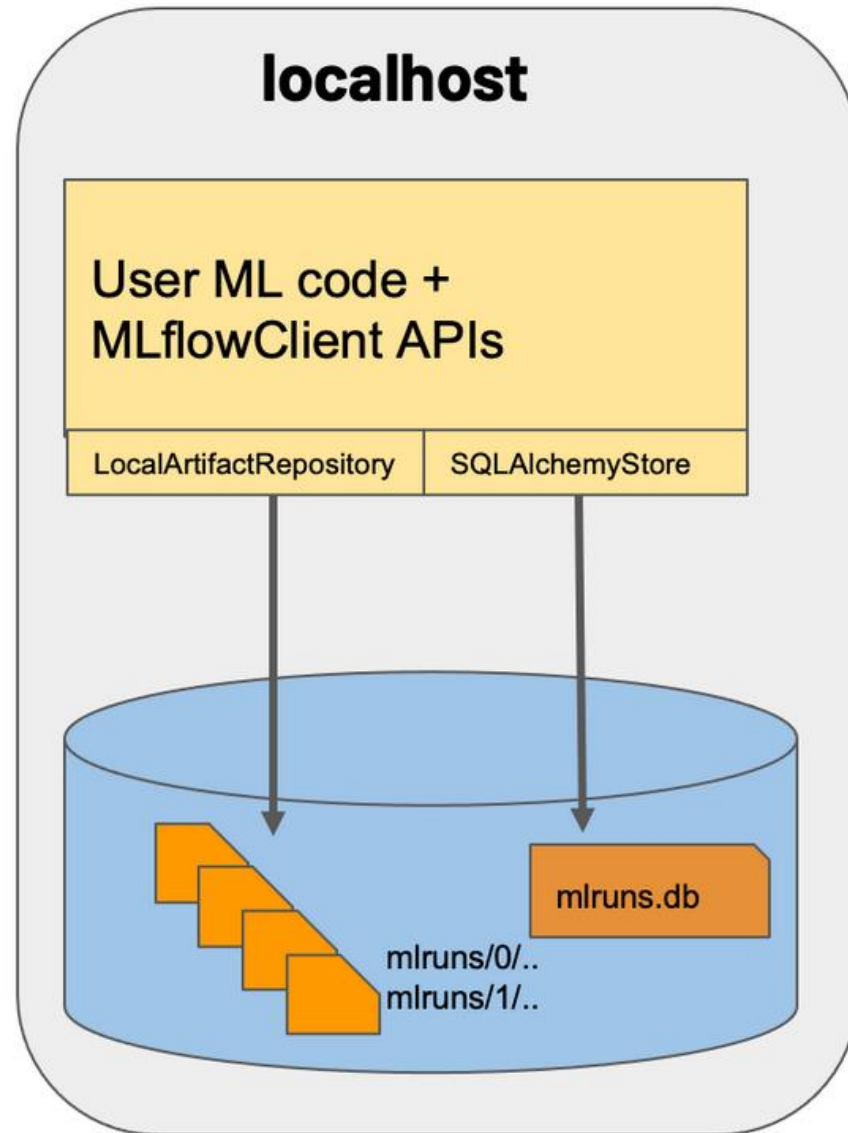


- Установка
 - `poetry add mlflow`
- Запуск сервиса
 - `mlflow server --backend-store-uri <PATH>`
 - `mlflow ui`



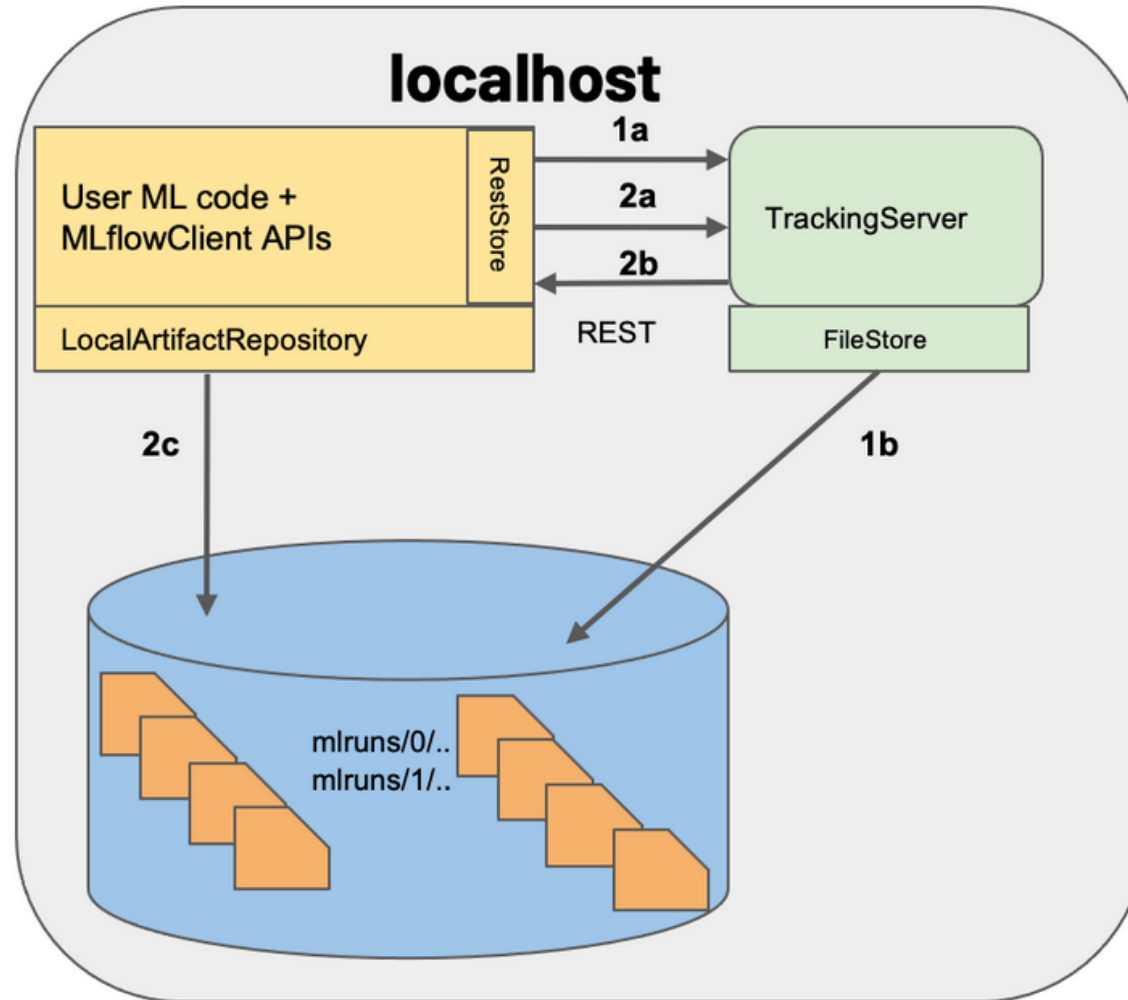


MLflow | Localhost with SQLite



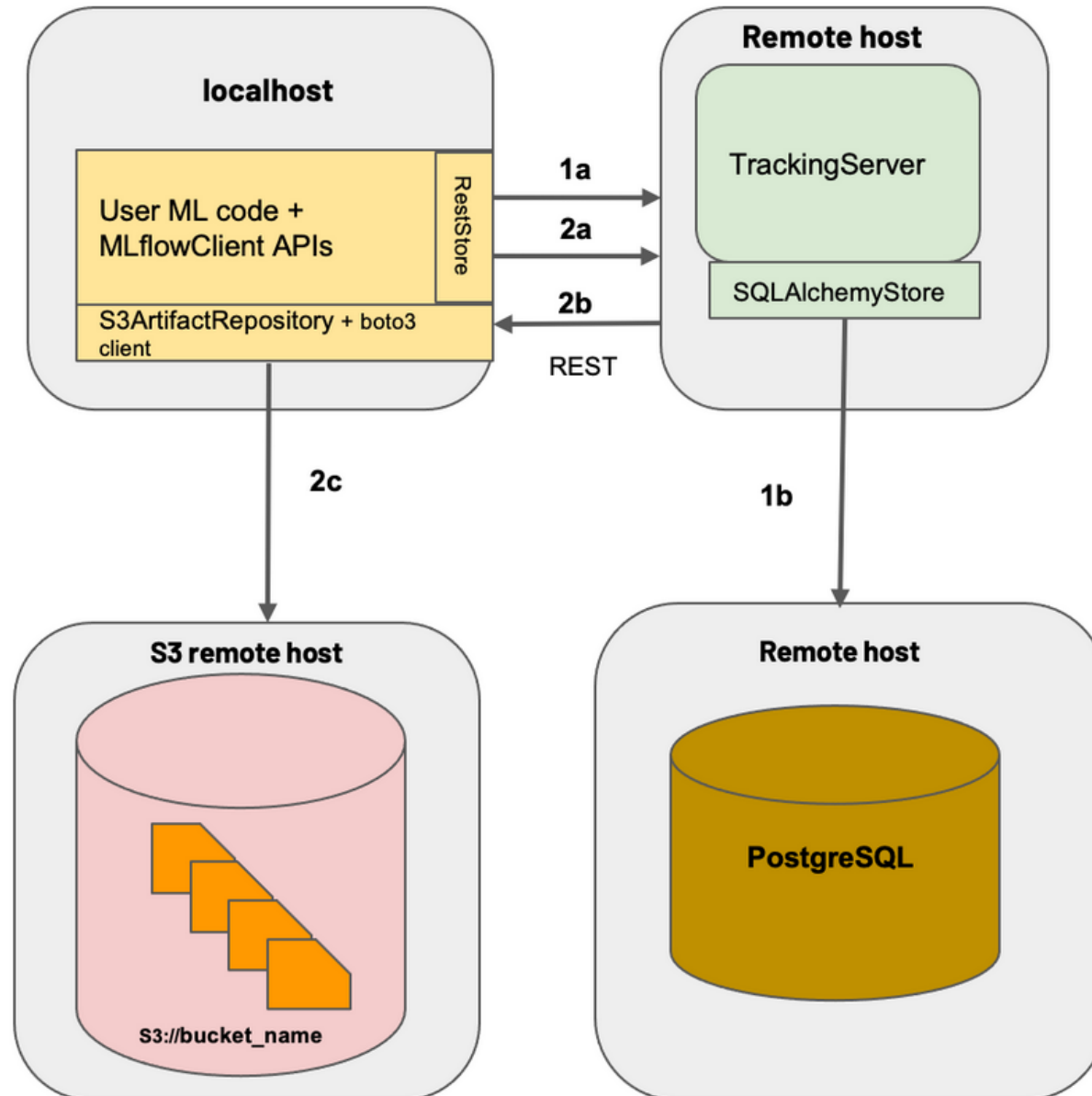


MLflow | Localhost with Tracking Server





MLflow | Tracking Server + backend + S3





Опыт развёртывания ML сервиса в production на примере проекта FindMyBike

Инструменты автоматизации ML исследований, DVC + MLFlow

Антон Ганичев
Газпромнефть ЦР
Руководитель направления NLP
t.me/pavel_kikin

