

# Advanced Topics and Use Cases

Lesson 8

---

DVC Tools for Data Scientists &  
Analysts

2021



# Course lessons

**Lesson 1.** Course Introduction

**Lesson 2.** Practices and Tools for Efficient Collaboration in ML projects

**Lesson 3.** Pipelines Automation and Configuration Management

**Lesson 4.** Versioning Data and Models

**Lesson 5.** Visualize Metrics & Compare Experiments with DVC and Studio

**Lesson 6.** Experiments Management and Collaboration

**Lesson 7.** Tools for Deep Learning Scenarios

**Lesson 8.** Advanced Topics and Use Cases



# Lesson Outline

- ◆ Data registry
- ◆ Run ML experiments in the cloud with CML
- ◆ Run experiments from UI?



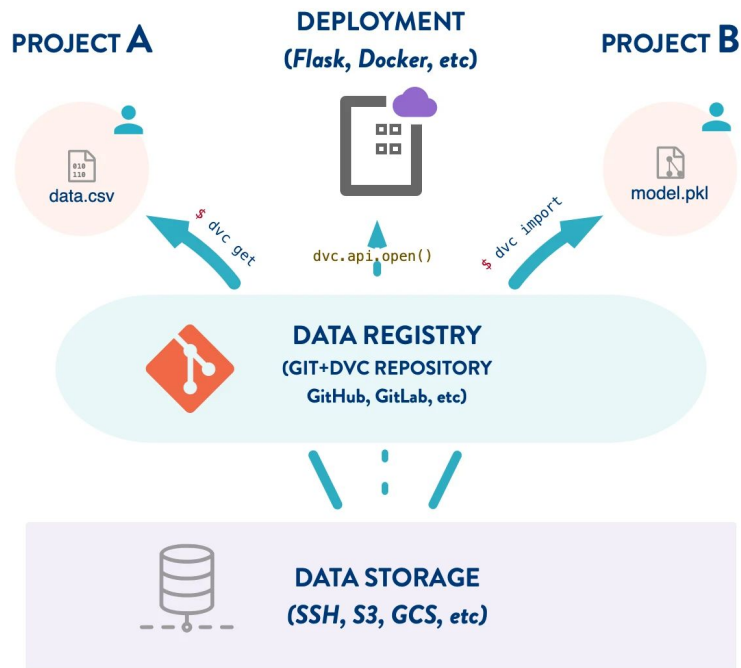


# Data Registry

# DVC tools for deep learning



- ◇ Reusability
- ◇ Persistence
- ◇ Storage optimization
- ◇ Data as code
- ◇ Security



# Commit successful experiment



Get/Import data  
from Data Registry

# CLI

```
dvc import <URL> dataset  
dvc get <URL> dataset
```

Open a model from  
Python code

# Python

```
import dvc.api.open  
  
model_path = 'model.pkl'  
repo_url = 'https://github.com/example/registry'  
  
with dvc.api.open(model_path, repo_url) as fd:  
    model = pickle.load(fd)  
  
# ... Use the model!
```



**Live code example**

# Example 1:

## Data Registry



**Run ML experiments in the cloud**  
**with CML**





## Live code example

### Example 2:

Run ML experiments in the  
cloud with CML and get  
automated reports



**Manage params and run experiments  
in Studio (with CML)**

# Compute needs to be automated



Change  
dataset  
version

Change  
hyperparameters

### Run new experiment

Input data files

c1fa36d9

data  
data.xml

params.yaml

3500

featureize  
max\_features

2

featureize  
ngrams

64

train  
min\_split

Commit changes

Update dataset & increase feature size

Optional description

☐ Commit directly to the master branch

☒ Create a new branch for this commit and start a pull request.

master-00071e8

Commit changes

Cancel

Type commit  
message

Run ML  
training



## Live code example

### Example 3:

Run ML experiments in the  
cloud with CML

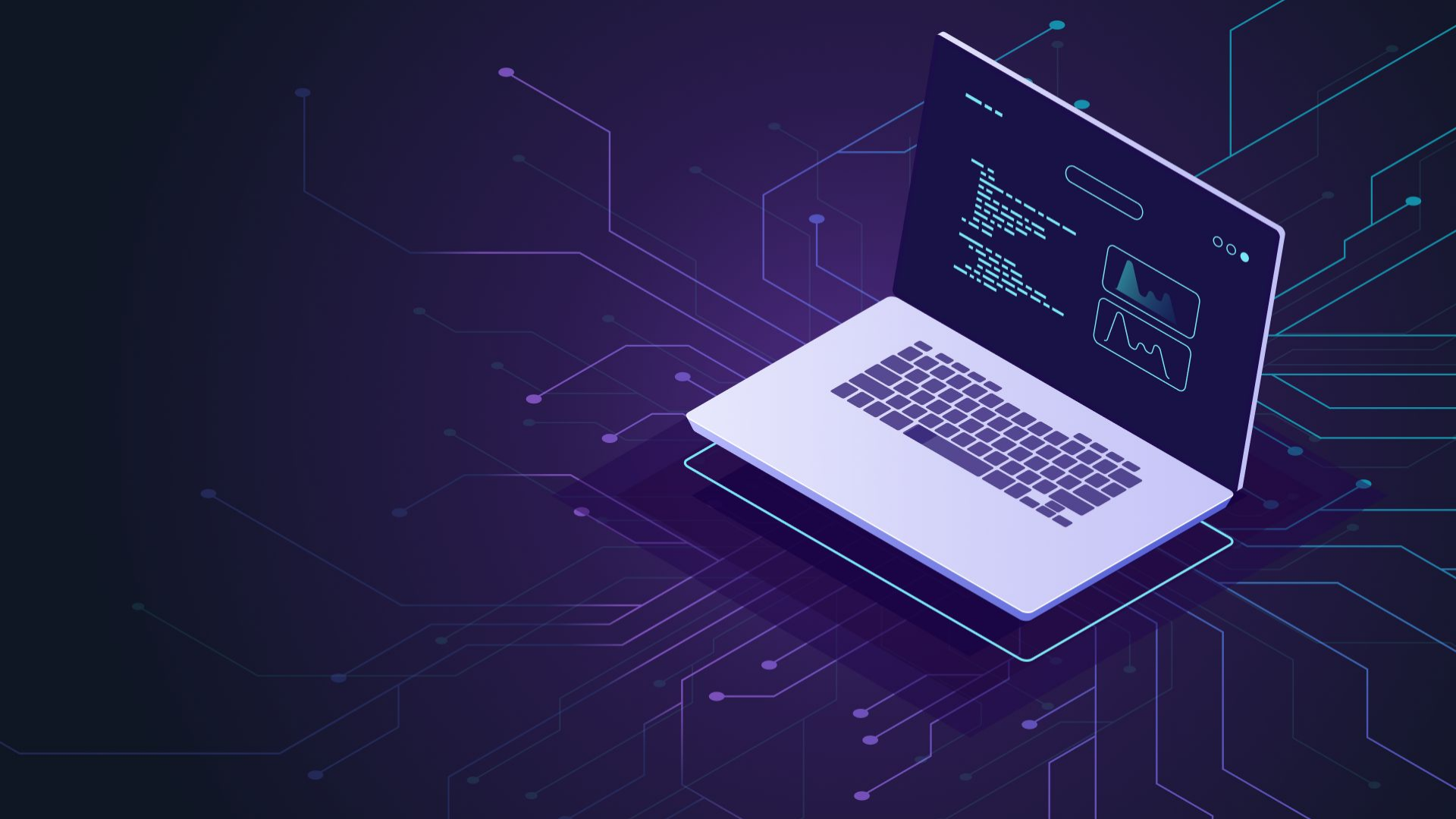


**What have we learned?**

# What have we learned?



1. Data registry
2. Run ML experiments in the cloud with CML
3. Run experiments from UI?





# Links



Data Science blueprint

<https://data-science-blueprint.readthedocs.io/en/latest/presentation/schema.html>