

Reinforcement Learning for RecSys – from Bandits to Offline RL with Ray RLlib

Jun Gong

Kourosh Hakhamaneshi

jungong@anyscale.com

kourosh@anyscale.com

\$whoami (Kourosh)

- RL Team Lead @Anyscale.
- Previously: PhD student at UC Berkeley working on RL in Robotics and design optimization

\$whoami (Jun)

- Senior Member of Engineering team @Anyscale.
- Previously: Engineer at Google and Facebook. Distributed RL at Loon.



Anyscale

Who we are: Original creators of Ray, a unified framework for scalable, distributed computing. Part of that framework are our libraries for ML and data processing.

What we do: Scalable compute for AI and Python

Why we do it: Scaling is a necessity, scaling is hard; make distributed computing easy and simple for all developers.



Some of RLlib's Industry Users



J.P.Morgan



TWO SIGMA



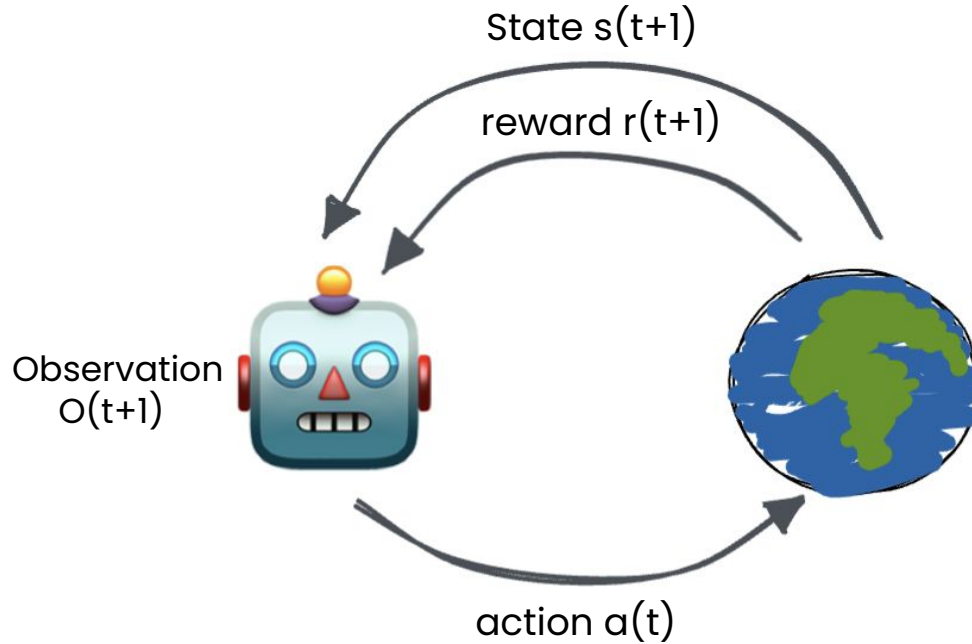
Overview of the tutorial

- Brief intro RL
- Brief intro RecSys
 - + Traditional Approaches
 - + Defining RecSys as an RL problem
- Online RL vs Offline RL
- Hands-on coding with python notebooks and scripts
- Thank you! Connect with us!

Goals – Understand:

- What are the advantages of using RL in RecSys?
- What are the pros and cons of offline RL in practical scenarios?

Brief intro RL

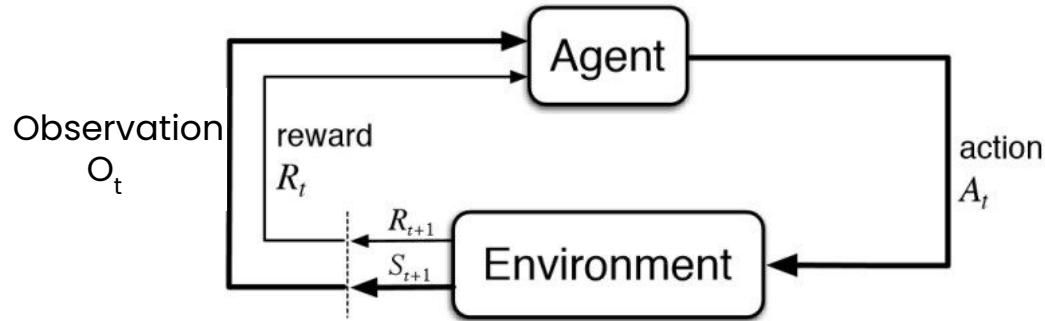


Algorithm

$$\max_{\pi} \mathbb{E}_{\pi} \left[\sum_t r(s_t, a_t) \right]$$



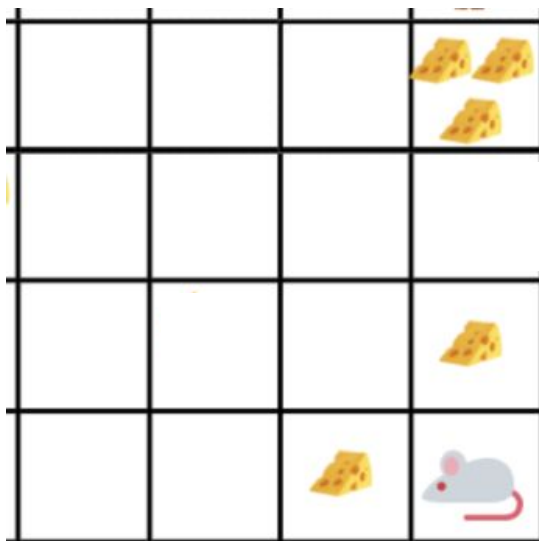
Brief intro RL formalization



$$(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$$

Discount factor γ in RL

- If $\gamma = 0$, the algorithm considers **1-step rewards only**.
- If $\gamma = 1$, the algorithm considers all future rewards equally.



Brief intro RecSys

Companies want to recommend content.



ML: Pointwise recommendations.



RL: Combine pointwise recommendations with session based data.



Traditional ML in RecSys

- Traditional ML (collaborative filtering) models are **static with respect to time**.
 - Ignores the **sequence of interactions** with a given user.
- Static models can be:
 - Too short-sighted and **miss out on Long-term, delayed rewards**
 - **Overlook important and changing user intents** or business conditions such as seasonality or promotional campaigns

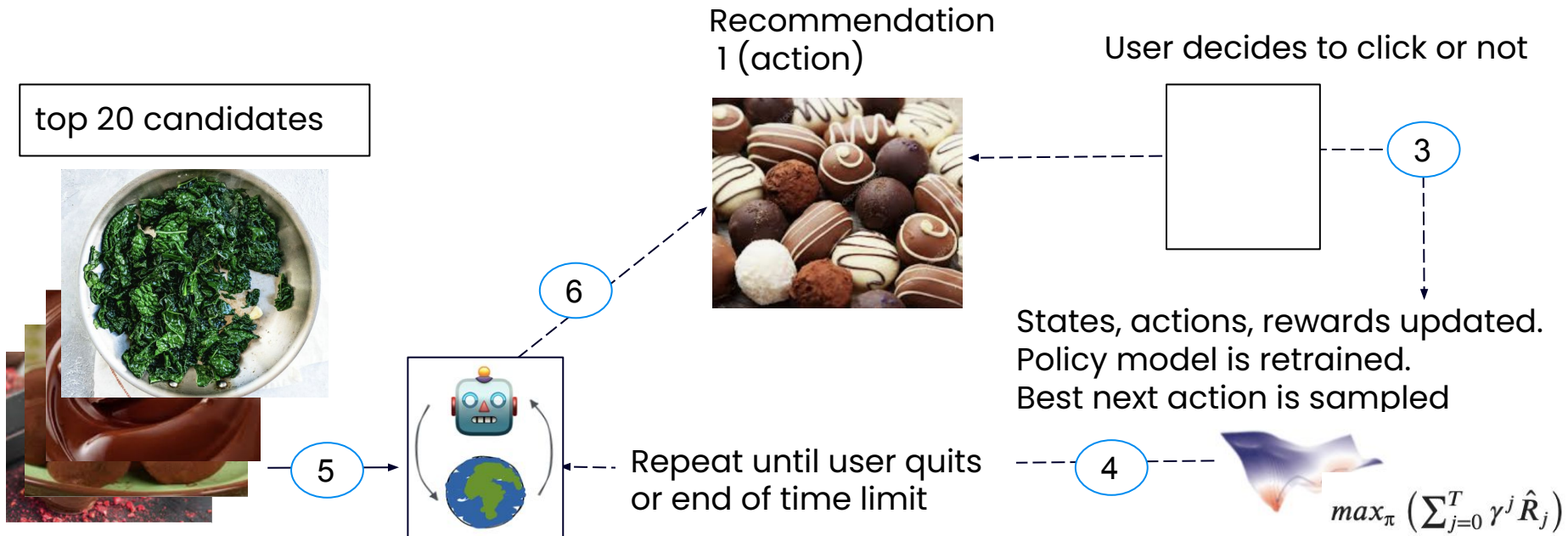
New way: RL in RecSys

- By formulating each user's session history as a sequence of decisions, **the RecSys problem can be converted into a sequential decision-making problem.**
 - **documents** = items to be recommended
 - **States** = previous item features, user features
 - **Actions** = recommended items
 - **Rewards** = long term satisfaction
 - **Gamma** = 0 (bandits) or 1 (RL)
 - **Agent** = Recommendation system
 - **Env** = Google's RecSim
 - **Algorithm** = RLlib algorithm



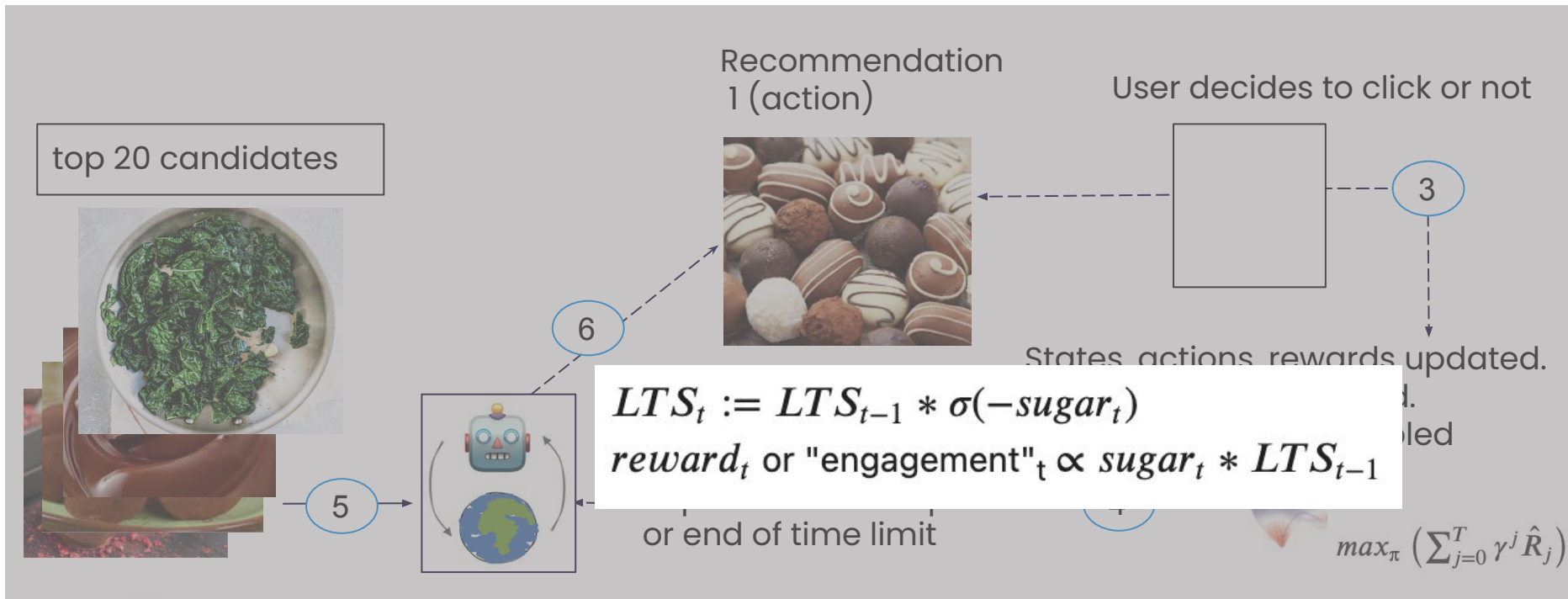
Environments Suitable for RL

Delayed Rewards & Long Term Satisfaction (LTS)

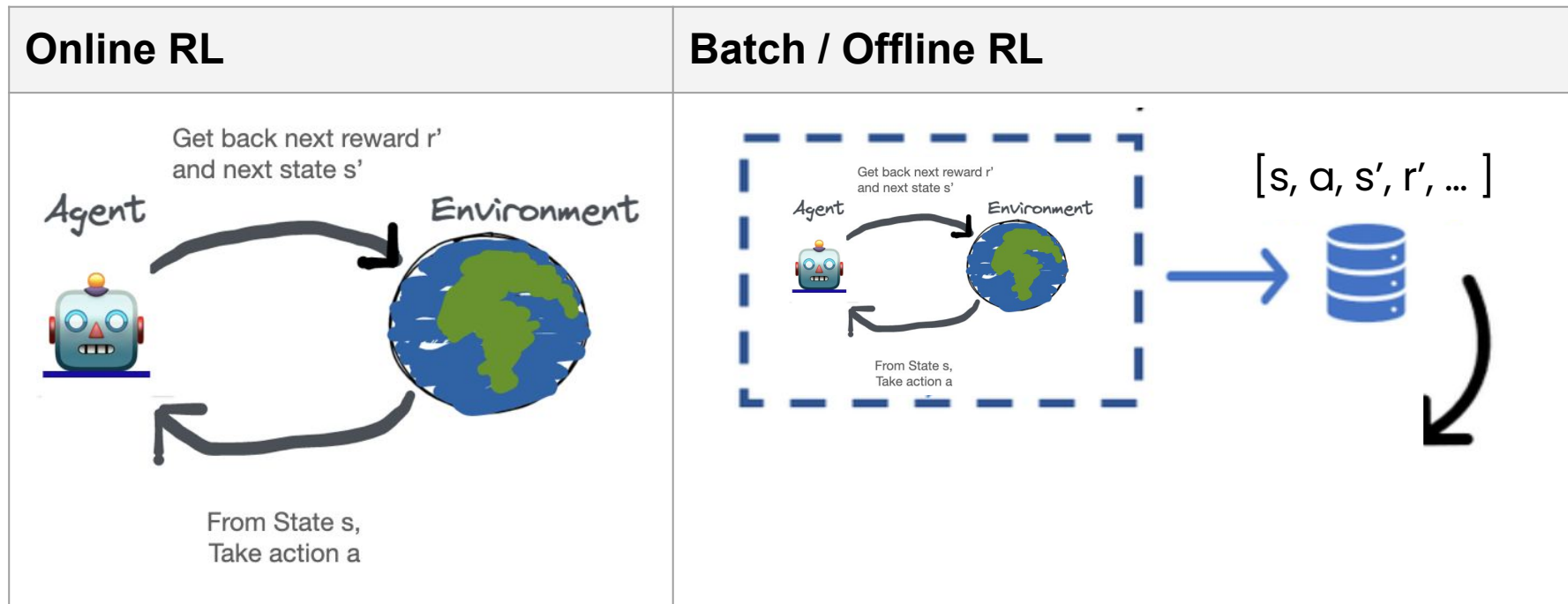


Environments Suitable for RL

Delayed Rewards & Long Term Satisfaction (LTS)



Online RL vs Offline RL



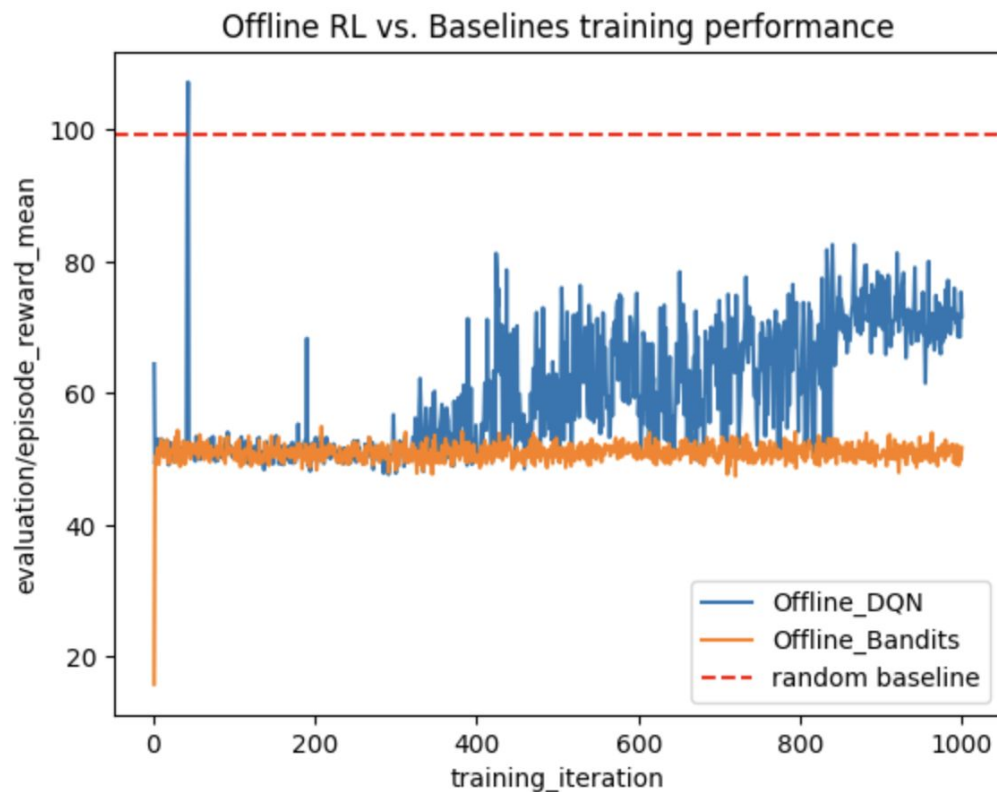
Today's Dataset

- RecSim environment with a random agent.
- Offline Bandit
- Offline RL



https://bit.ly/rllib_odsc_west_2022

Sample result from notebook



Your Anyscale Cluster

1. Claim username/password at https://bit.ly/rllib_odsc_west_2022
 - a. Update the “**Status**” column to “**Claimed**”
2. Go to Console: <http://console.anyscale.com/>
3. Enter username (for the email) and password



Scale your application from
your laptop to the cloud

Get started

Work email

1

john@acme.com

2

3

Next



Your Anyscale Cluster

anyscale

- Home
- Projects
- Interactive sessions
- Jobs
- Services
- Clusters**
- Configurations

Clusters

+ Create **Start** **Terminate** **Archive**

Search names Cluster status Created by Include archived

<input type="checkbox"/>	Name	Status ↓	Active resources	Cost ? ↑↓	Cluster environment
<input type="checkbox"/>	cluster-0	Terminated	None	\$0.80	ray_tutorial_app_config_allentest200:1

4 5

Your Anyscale Cluster

The screenshot displays the Anyscale Ray console interface. On the left is a dark blue sidebar with navigation links: **anyscale**, Home, Projects (highlighted), Interactive sessions, Jobs, Services, Clusters, and Configurations. The main content area shows the breadcrumb **Ray-Tutorial > cluster-0**. An orange arrow points to the **Jupyter** button, with the word **click** in orange text next to it. Other buttons include **Dashboard**, **Grafana**, and **Terminate**.

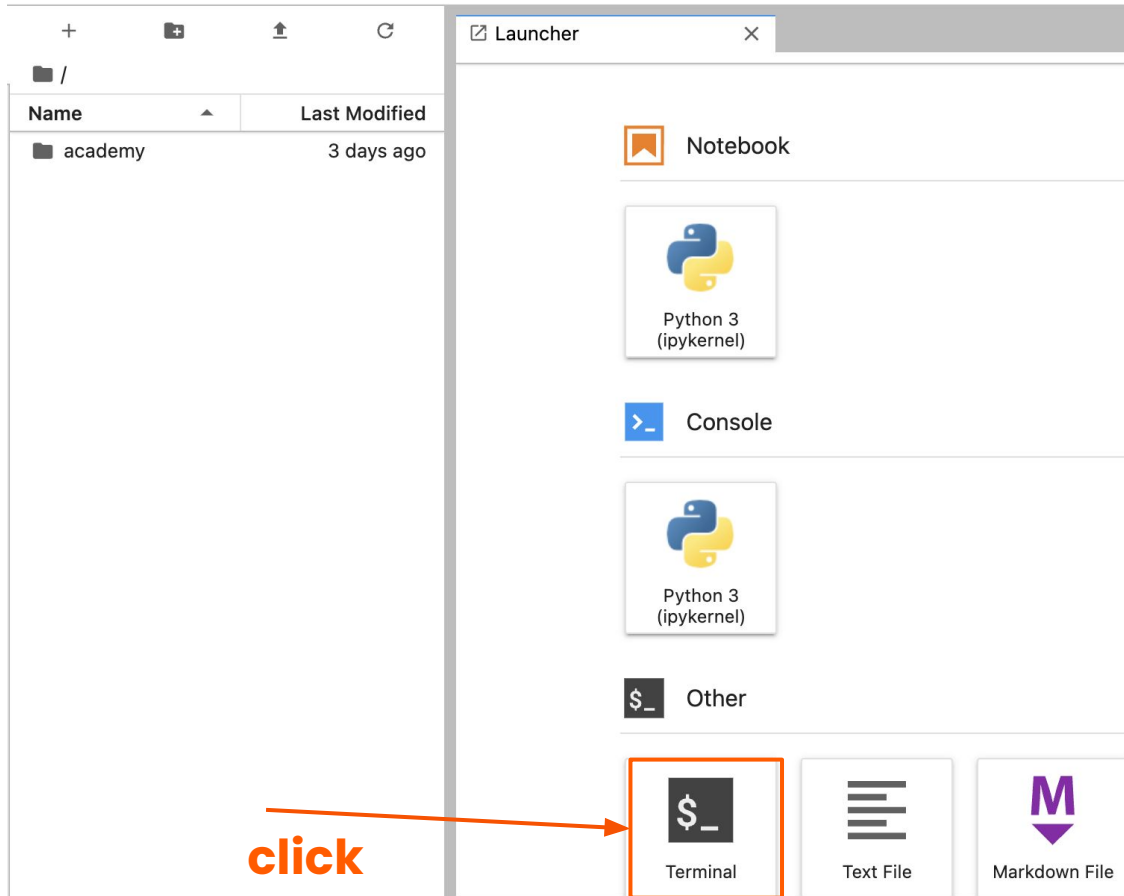
About this cluster

Status	ID	Created by
Active (auto-suspend in 2880 minutes)	ses_QSpdDJDjX3pu4Xz93iD9Sb7p	yinhaonan55+200@gmail.com
Created at	Access	Project
Jul 18, 2022, 2:13:41 PM	Only admins and you can view and edit	Ray-Tutorial

Resource usage

CPU	Object store memory	GPU
0 utilized / 8 running	0 B utilized / 6.87 GiB running	-
Cost since last start	Cost since creation	
\$0.80	\$0.80	

Your Anyscale Cluster



The screenshot shows the Anyscale Launcher interface. On the left is a file explorer with a table of files:

Name	Last Modified
academy	3 days ago

On the right is the main workspace area. It has a tab labeled 'Launcher'. Below the tab, there are several options:

- Notebook
- Python 3 (ipykernel)
- Console
- Python 3 (ipykernel)
- Other
- Terminal (highlighted with an orange box and an arrow pointing to it with the word 'click')
- Text File
- Markdown File



ray@ip-10-0-104-35: ~/Ray X

```
(base) ray@ip-10-0-104-35:~/Ray-Tutorial$ ls
```

academy

```
(base) ray@ip-10-0-104-35:~/Ray-Tutorial$ cd academy
```

```
(base) ray@ip-10-0-104-35:~/Ray-Tutorial/academy$ git pull
```

```
remote: Enumerating objects: 121, done.
```

```
remote: Counting objects: 100% (121/121), done.
```

```
remote: Compressing objects: 100% (78/78), done.
```

```
remote: Total 108 (delta 53), reused 70 (delta 29), pack-reused 0
```

```
Receiving objects: 100% (108/108), 4.76 MiB | 4.13 MiB/s, done.
```

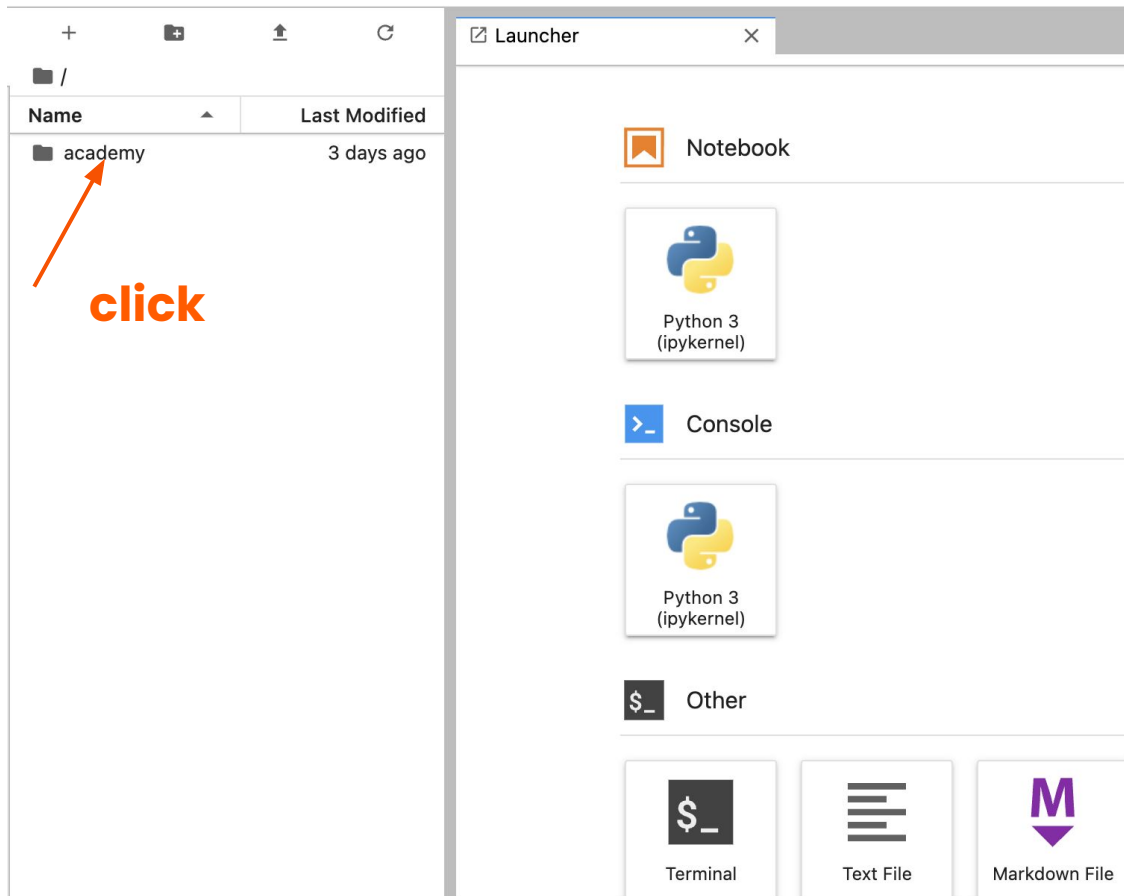
```
Resolving deltas: 100% (53/53), completed with 8 local objects.
```

```
From https://github.com/anyscale/academy
```

```
    a5ee457..405730b  main      -> origin/main
```



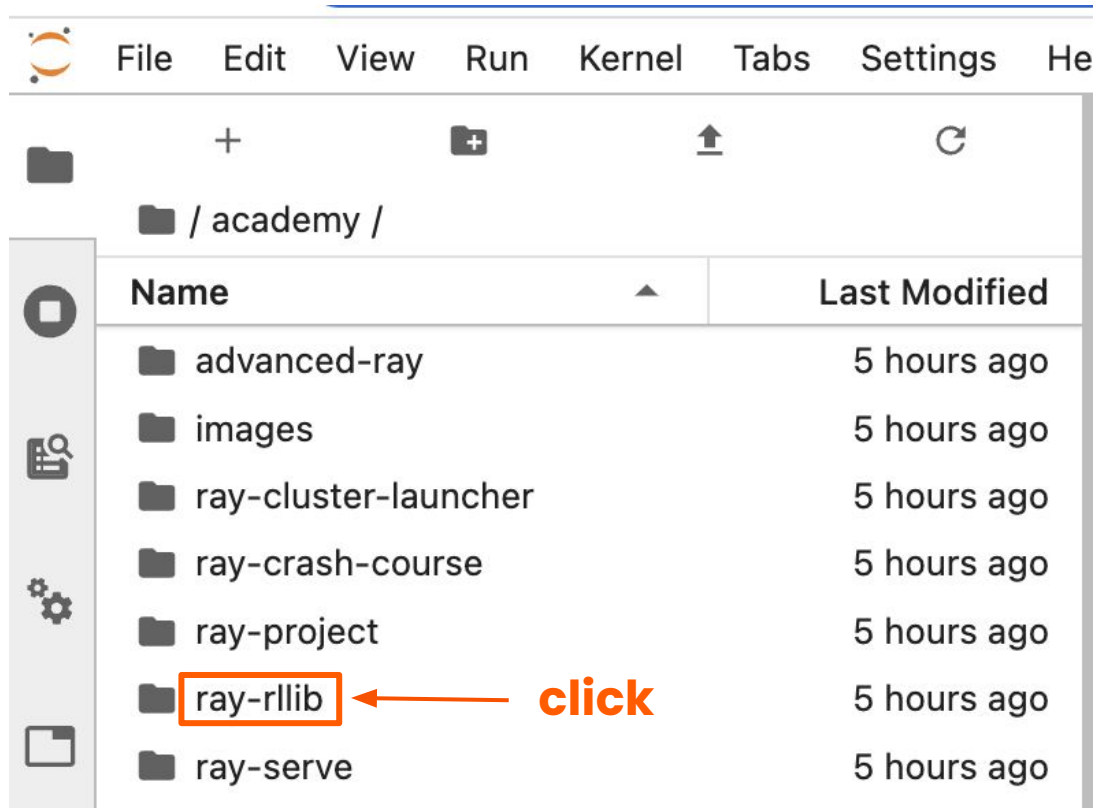
Your Anyscale Cluster



The screenshot shows the Anyscale Launcher interface. On the left, a file explorer displays a folder named 'academy' under the root directory '/'. The folder is highlighted with a red arrow and the word 'click' in orange text. The file explorer has columns for 'Name' and 'Last Modified', with 'academy' listed as '3 days ago'. On the right, the 'Launcher' window shows a sidebar with options: 'Notebook', 'Console', and 'Other'. The 'Notebook' section is active, displaying two 'Python 3 (ipykernel)' options. The 'Other' section at the bottom shows 'Terminal', 'Text File', and 'Markdown File' options.



Your Anyscale Cluster

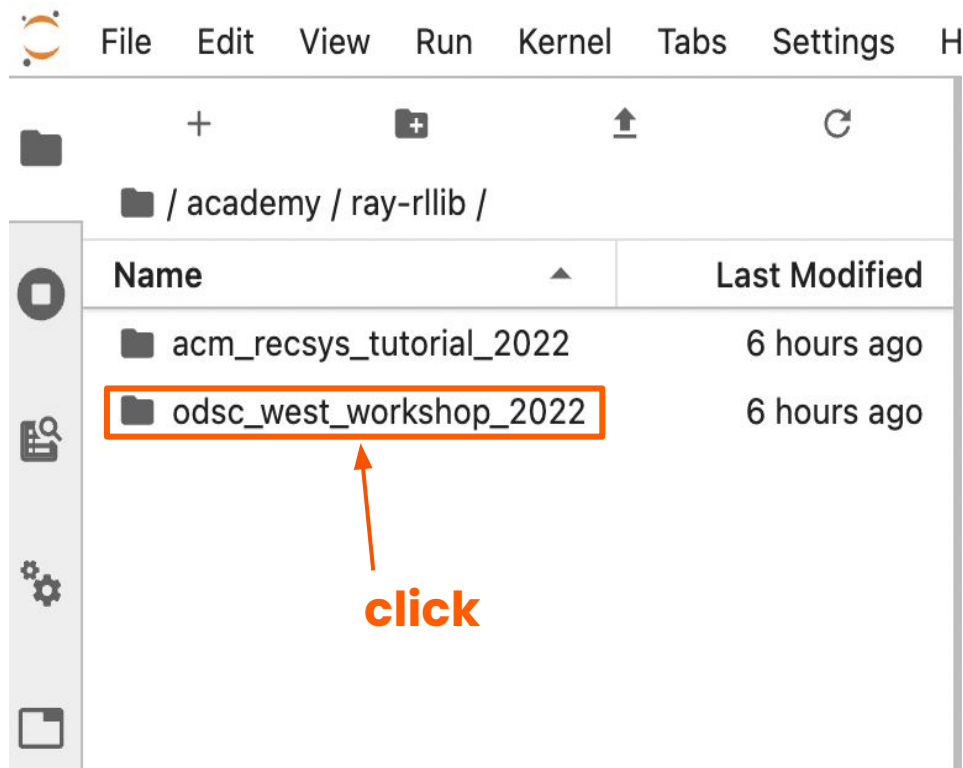


The screenshot shows the Anyscale IDE interface. At the top is a menu bar with options: File, Edit, View, Run, Kernel, Tabs, Settings, and Help. Below the menu bar is a toolbar with icons for creating a new file, creating a new folder, uploading a file, and refreshing. The main area displays a file explorer for the path '/ academy /'. A table lists the contents of this directory:

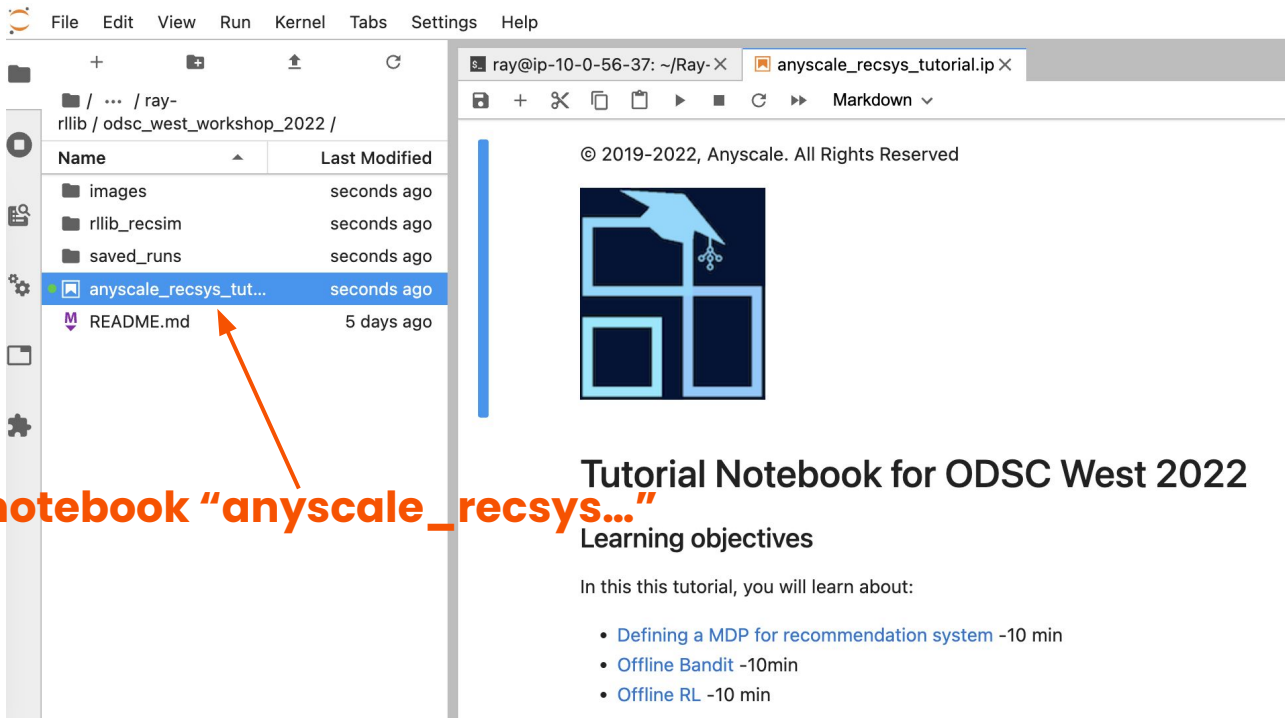
Name	Last Modified
advanced-ray	5 hours ago
images	5 hours ago
ray-cluster-launcher	5 hours ago
ray-crash-course	5 hours ago
ray-project	5 hours ago
ray-rlib	5 hours ago
ray-serve	5 hours ago

An orange box highlights the 'ray-rlib' folder, and an orange arrow points to it with the word 'click'.

Your Anyscale Cluster



Your Anyscale Cluster



File Edit View Run Kernel Tabs Settings Help

ray@ip-10-0-56-37: ~/Ray- X anyscale_recsys_tutorial.ip X

© 2019-2022, Anyscale. All Rights Reserved

Tutorial Notebook for ODSC West 2022

Learning objectives

In this this tutorial, you will learn about:

- [Defining a MDP for recommendation system](#) -10 min
- [Offline Bandit](#) -10min
- [Offline RL](#) -10 min

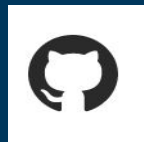
Open notebook "anyscale_recsys..."

Thank you.

We would love to connect with you!



Twitter – @anyscalecompute | @raydistributed



GitHub – <https://github.com/ray-project/ray>



Slack – <https://www.ray.io/community>



Discuss – <https://discuss.ray.io/>