

# From your laptop to all resource that you need

@PyDataVenice #PyDataGlobal #2023



# Alessandra Bilardi

Data & Automation Specialist

- Coderdojo member
- PyData Venezia member
- AWS User Group Venezia member

✉ [alessandra.bilardi@gmail.com](mailto:alessandra.bilardi@gmail.com)

 [@abilardi](https://twitter.com/abilardi)

 [bilardi](https://www.linkedin.com/in/bilardi)

# Alessandra Bilardi

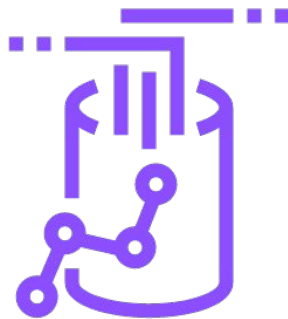
Data & Automation Specialist

- Coderdojo member
- PyData Venezia member
- AWS User Group Venezia member

✉ [alessandra.bilardi@gmail.com](mailto:alessandra.bilardi@gmail.com)

🐦 [@abilardi](https://twitter.com/abilardi)

🌐 [bilardi](https://www.linkedin.com/in/bilardi)



ingestion



analysis



prediction



visualization

---

# Agenda

Goal

Solution

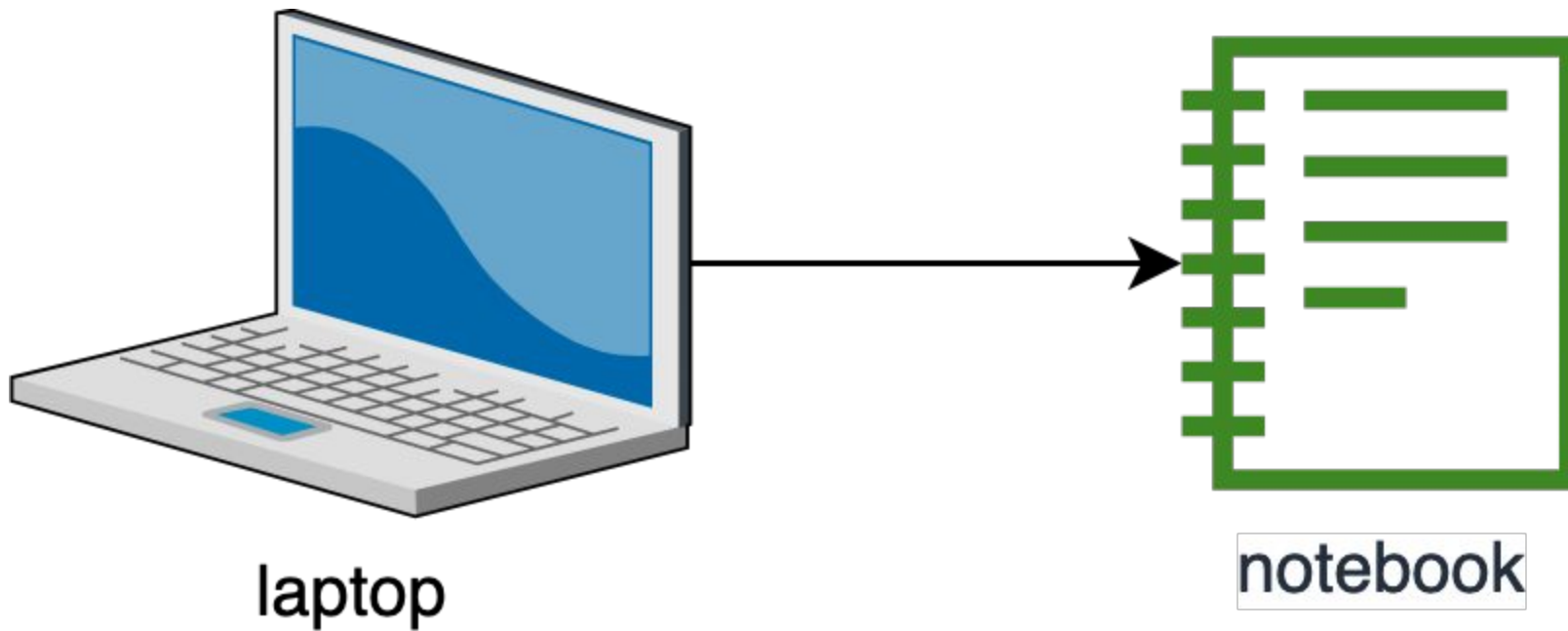
Cloud

Demo

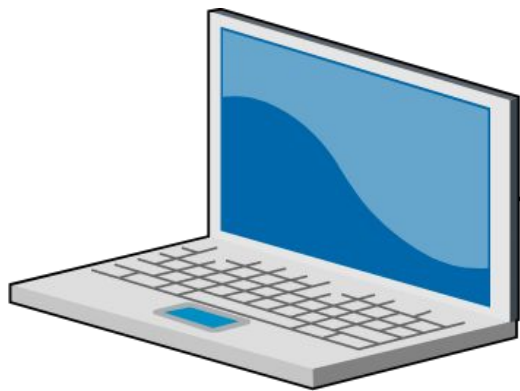
Best practices

# Goal

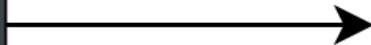
---



# Goal



laptop

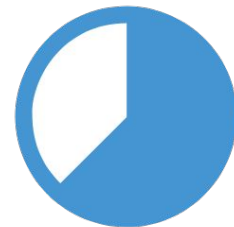


notebook

performance

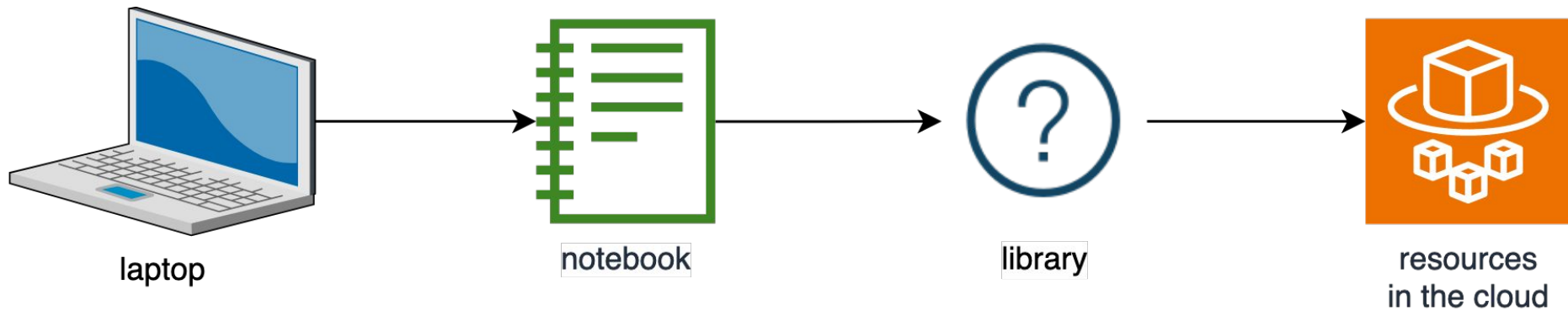


timing



resources

# Solution



—

may

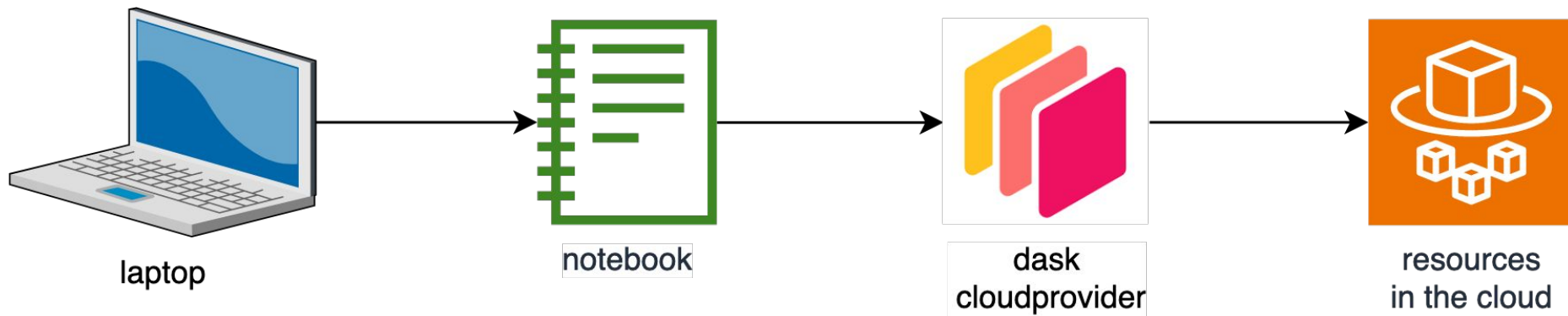
open source

with you



# Solution

---





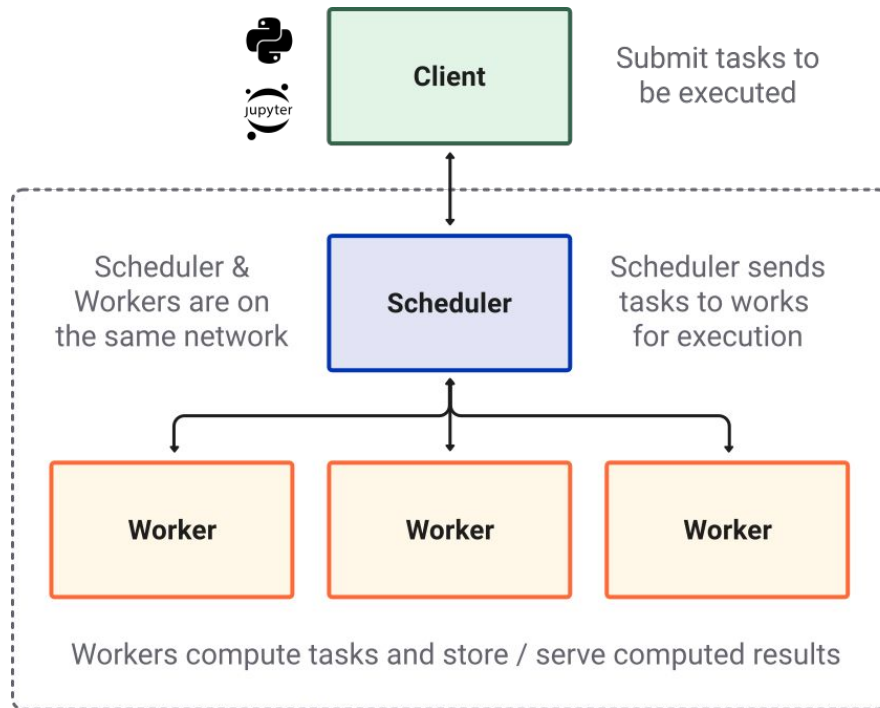
# Dask

Fundamentals

ML integrations

- Dataframe
  - Array
  - Bag
  - Delayed
  - Futures
  - Distributed
- 
- Sklearn & Joblib
  - XGBoost & LightGBM
  - PyTorch
  - Keras and Tensorflow

# Dask Cluster



—

scale

the python tools

you love

# Cloud

---

# Top 10 Cloud Provider 2023



# Dask Cloud Provider


[Dask](#)
[Distributed](#)
[Dask ML](#)
[Examples](#)
[Ecosystem](#)
[Community](#)

Dask Cloud Provider  
0+untagged.50.gef21317  
documentation

## OVERVIEW

[Installation](#)
[Configuration](#)

## PROVIDERS

[Amazon Web Services \(AWS\)](#)
[DigitalOcean](#)
[Google Cloud Platform](#)
[Microsoft Azure](#)
[Hetzner](#)

## ADVANCED



## Dask Cloud Provider

*Native Cloud integration for Dask.*

This package provides classes for constructing and managing ephemeral Dask clusters on various cloud platforms.

Dask Cloud Provider is one of many options for deploying Dask clusters, see [Deploying Dask](#) in the Dask documentation for an overview of additional options.

To use a cloud provider cluster manager you can import it and instantiate it. Instantiating the class will result in cloud resources being created for you.

```
from dask_cloudprovider.aws import FargateCluster
cluster = FargateCluster(
    # Cluster manager specific config kwargs
)
```

You can then construct a Dask client with that cluster object to use the cluster.

```
from dask.distributed import Client
client = Client(cluster)
```





# Today

## Goal:

to run our code on  
resources in the cloud

## Solution:

to use *AWS*  
and dask-cloudprovider

- our code with dask
  - for local
- configuration of tools
- our code modified
  - for cloud



# Code

---

# Configuration

---



# Configuration

AWS account:  
creation & configuration

Laptop side:  
configure AWS  
use AWS in our code

- AWS configuration
  - [create account](#)
  - [create credentials](#)
- on our laptop
  - [install aws-cli](#)
  - [configure credentials](#)
- in our code using libraries
  - boto3
  - [dask-cloudprovider](#)

# Code

---



# Results

## Goal:

to run our code on  
resources in the cloud

## Solution:

use AWS  
and dask-cloudprovider

- on Laptop
  - workers: 2
  - iters: 500
  - time: 271 seconds
- on AWS, in the cloud
  - workers: 4
  - iters: 500
  - time: 84 seconds



# Best practices

AWS console access:  
don't use root user !

Dask cluster:  
delete the cluster !

- AWS
  - use multi-factor authentication (MFA)
  - create your user with only the permissions that you need
- Laptop
  - update AWS credentials
  - create an environment
- Dask
  - use only the resources that you need
  - remember to delete the cluster for saving costs

—

may

data power

with you

# Links

---

## Support

- AWS getting started, <https://docs.aws.amazon.com/accounts/latest/reference/welcome-first-time-user.html>
- AWS free support, <https://discord.gg/N97snpCQ>
- Dask Overview, [https://tutorial.dask.org/00\\_overview.html](https://tutorial.dask.org/00_overview.html)
- Dask Cloud Provider, <https://cloudprovider.dask.org/en/latest/>

## References

- <https://ml.dask.org/joblib.html>
- <https://gist.github.com/jacobtomlinson/ee5ba79228e42bcc9975faf0179c3d1a>





# Thanks for listening.

@PyDataVenice #PyDataGlobal #2023