



Data science at scale at AWS

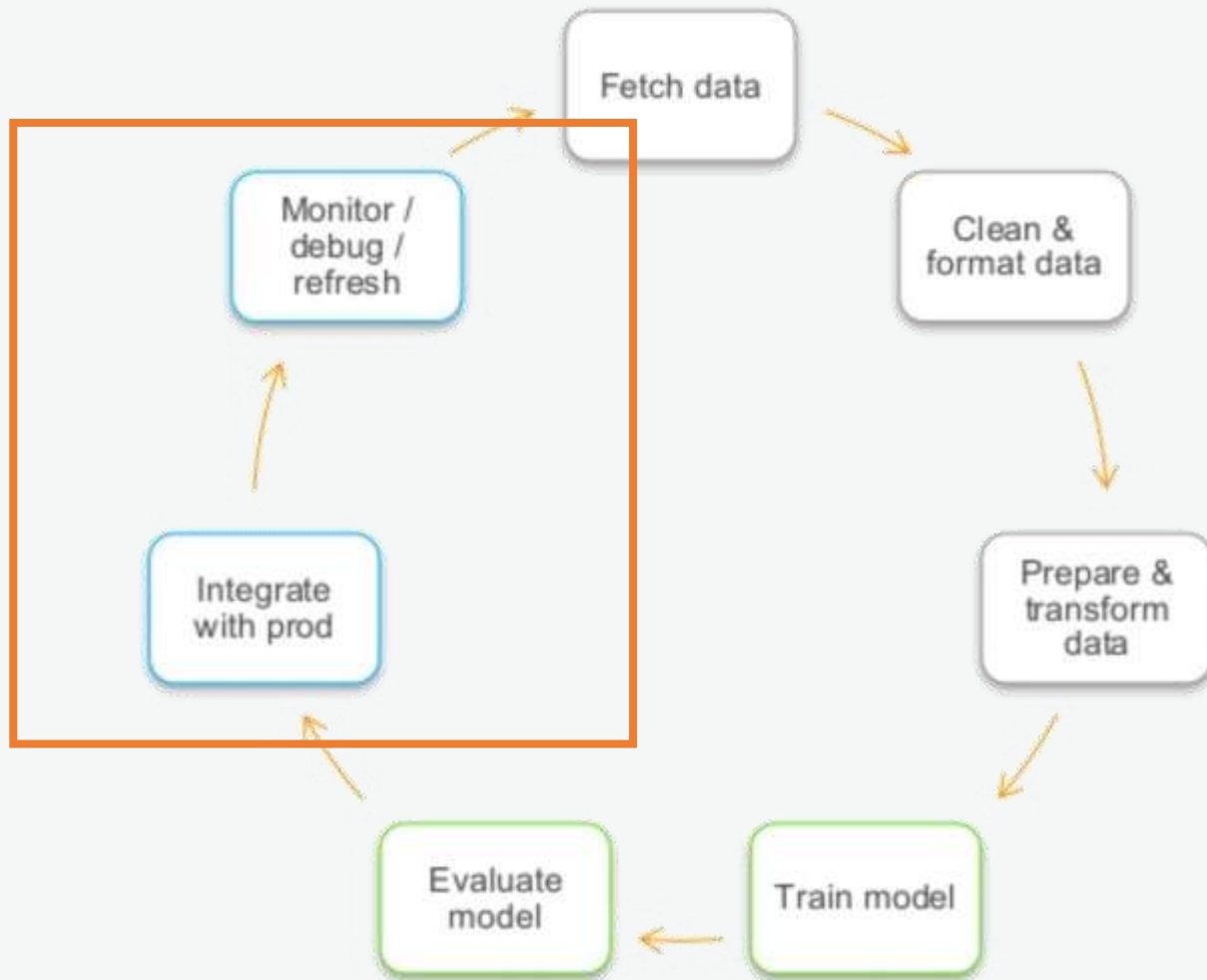
Dzidas Martinaitis



How does data science help our customers?

- Identify the customers who need help onboarding to our platform
- Spot unsatisfied customers in a timely manner
- Identify fraudulent activity
- Guide the business leaders through the most optimal path

Stages of a Data Science project



Why?

- It is not my job!
- Scaling

Passing the project



Benefits of owning the production environment

Code review

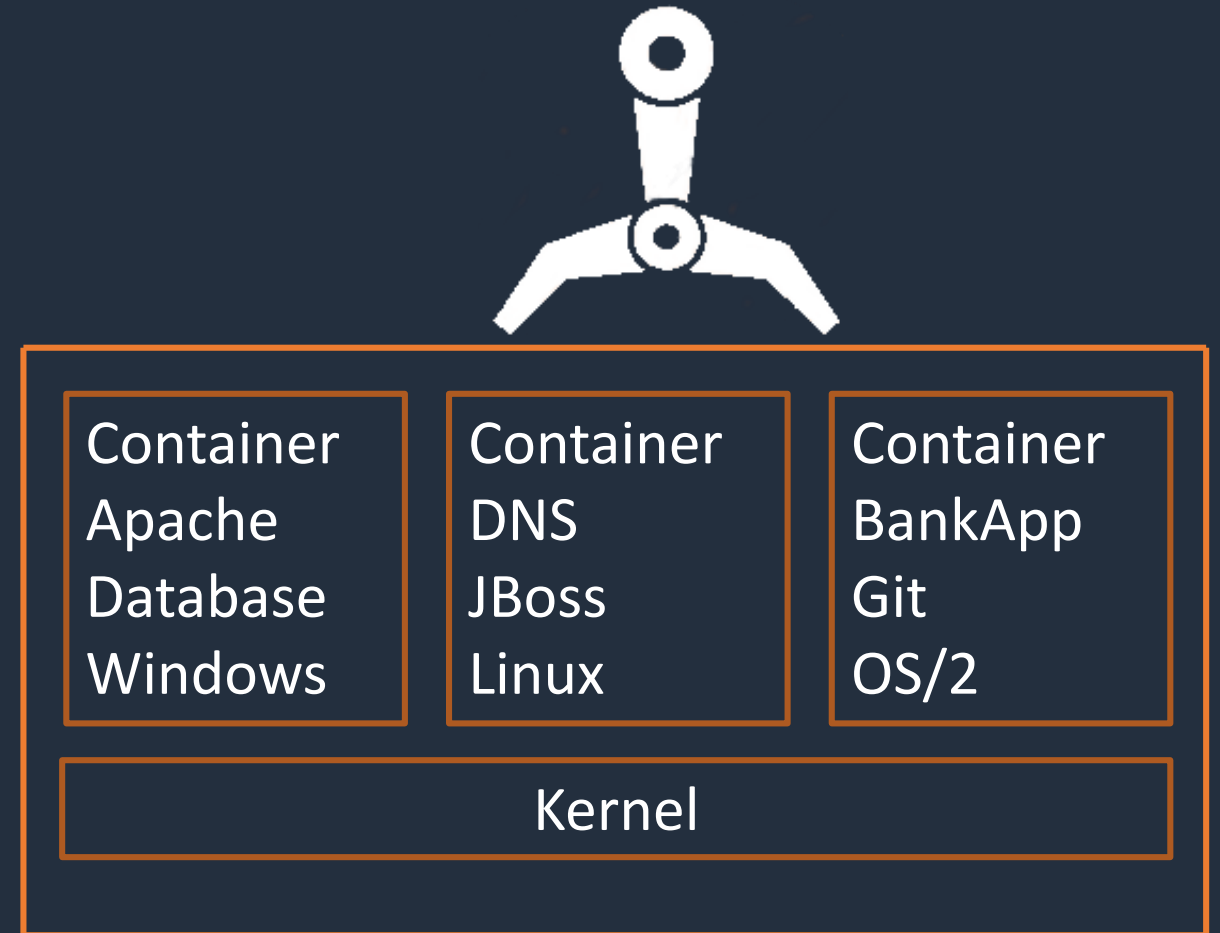
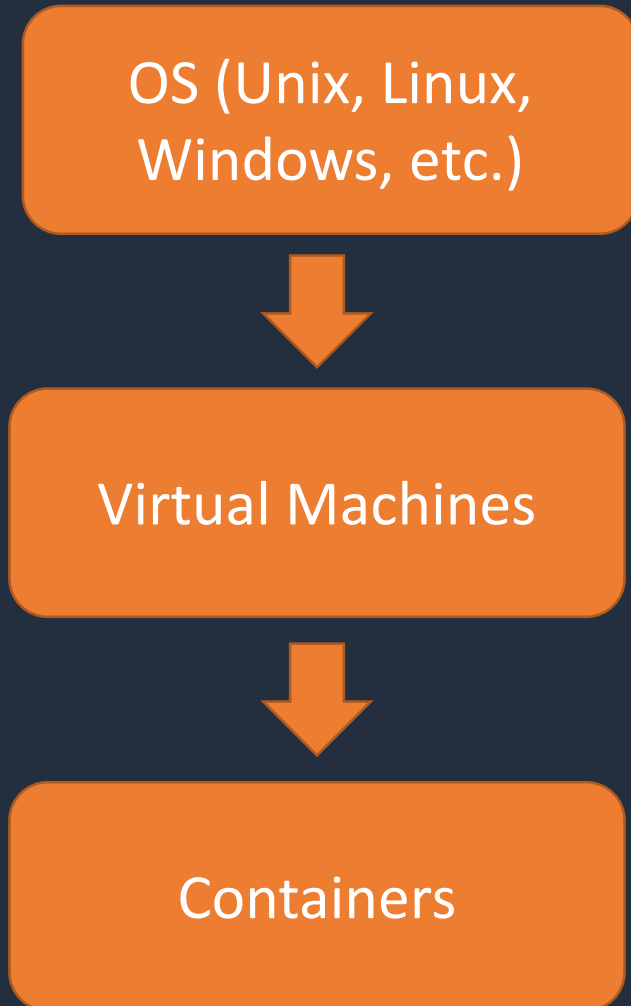
Model monitoring, refresh
Comparison of the models (A/B testing)

Scalability
Decoupling, model as a service
Different development stacks (R, Python, etc.)

Ownership – we own the project



What is a container?



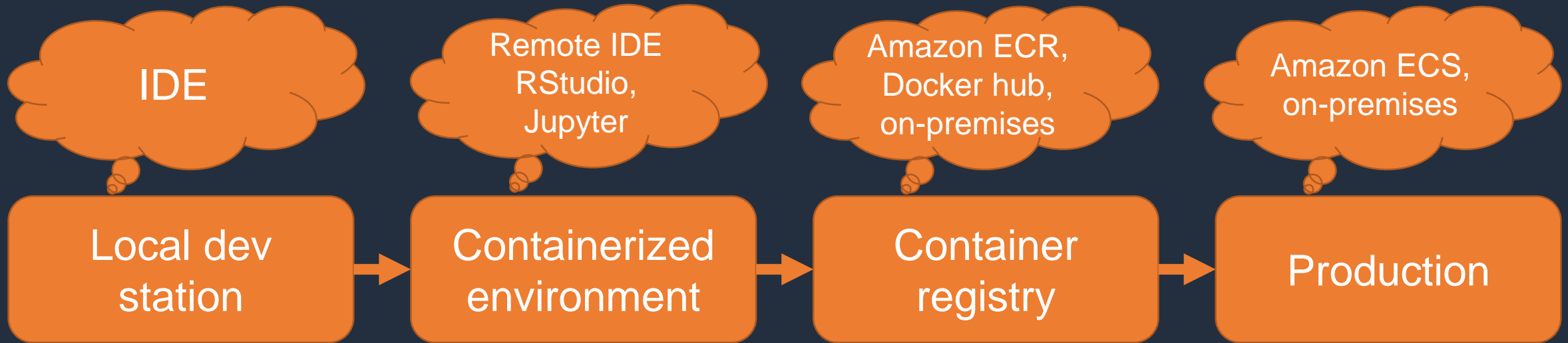
Dockerfile

```
FROM ubuntu:18.04
ENV DEBIAN_FRONTEND noninteractive

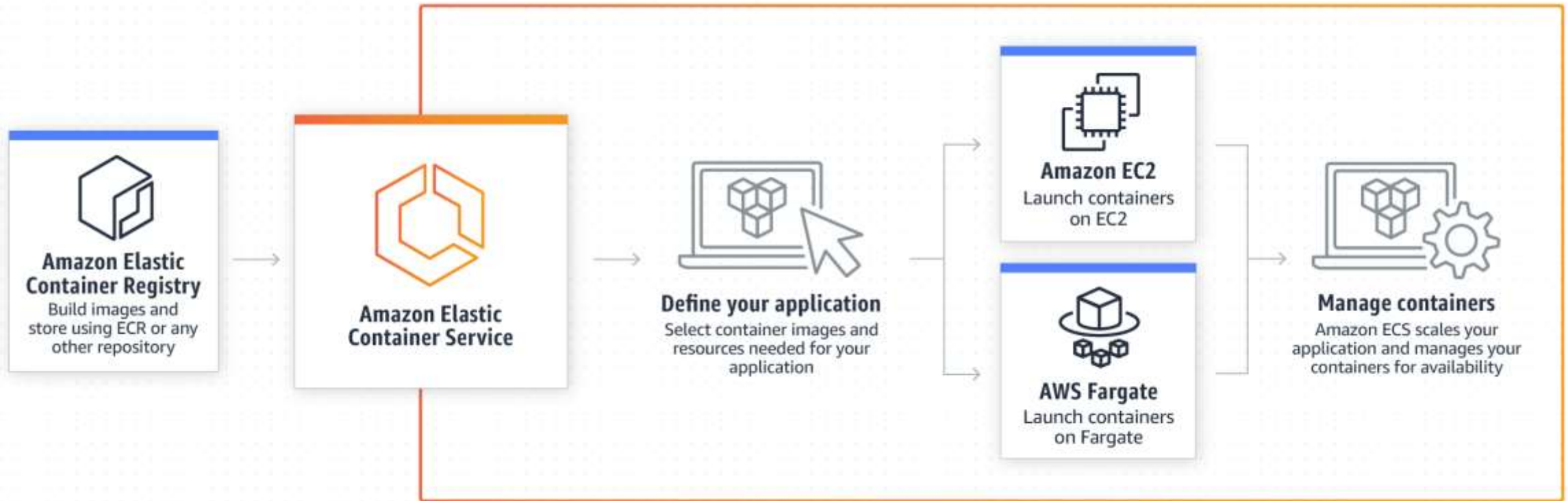
RUN apt-get update \
    && apt-get install -qy --no-install-recommends \
        wget \
        r-base \
        r-base-dev \
        r-recommended \
        ca-certificates

RUN R -e "install.packages(c('plyr', 'dplyr', 'RJDBC', 'mda', 'plumber') ,
    repos='https://cloud.r-project.org')"
```

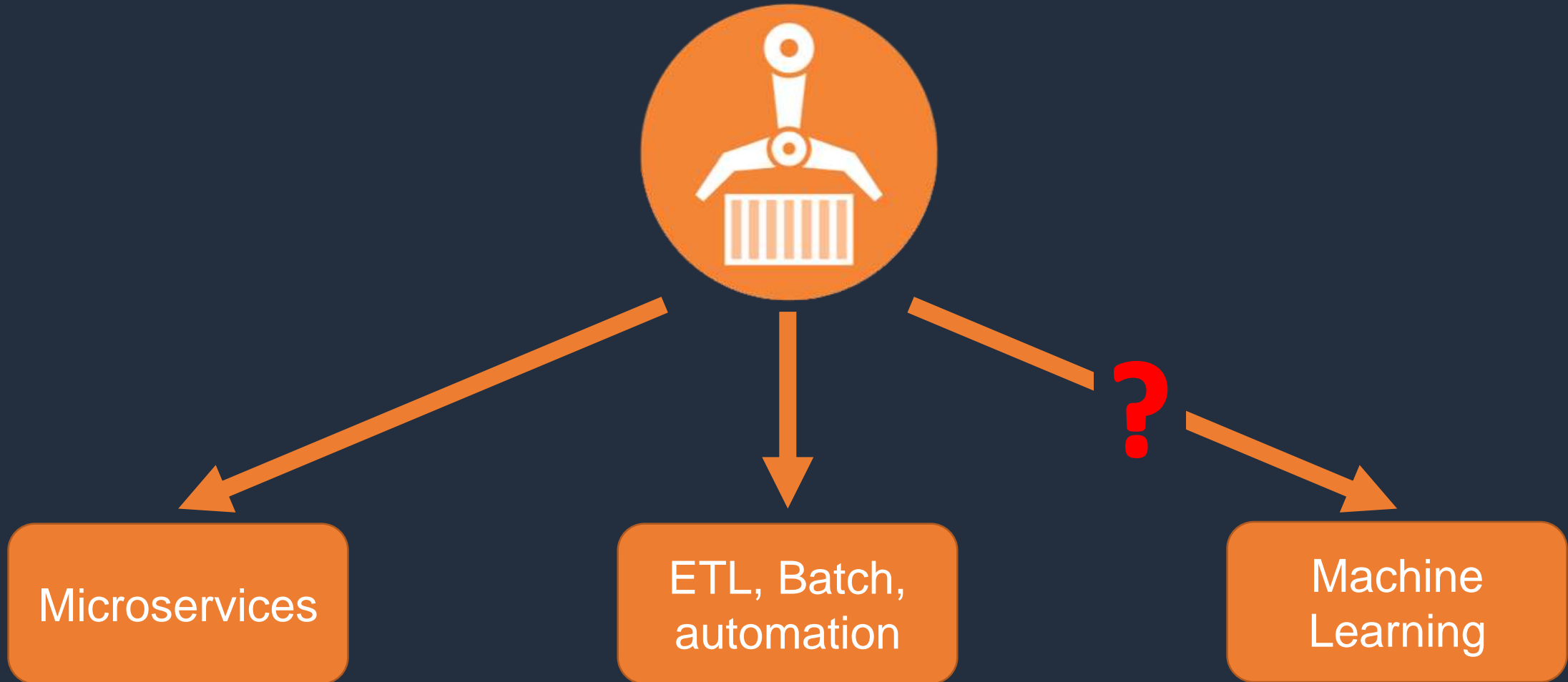
How to use Containers?



Amazon Elastic Container service



Is containerization the best solution?



Training, deploying and scaling the models

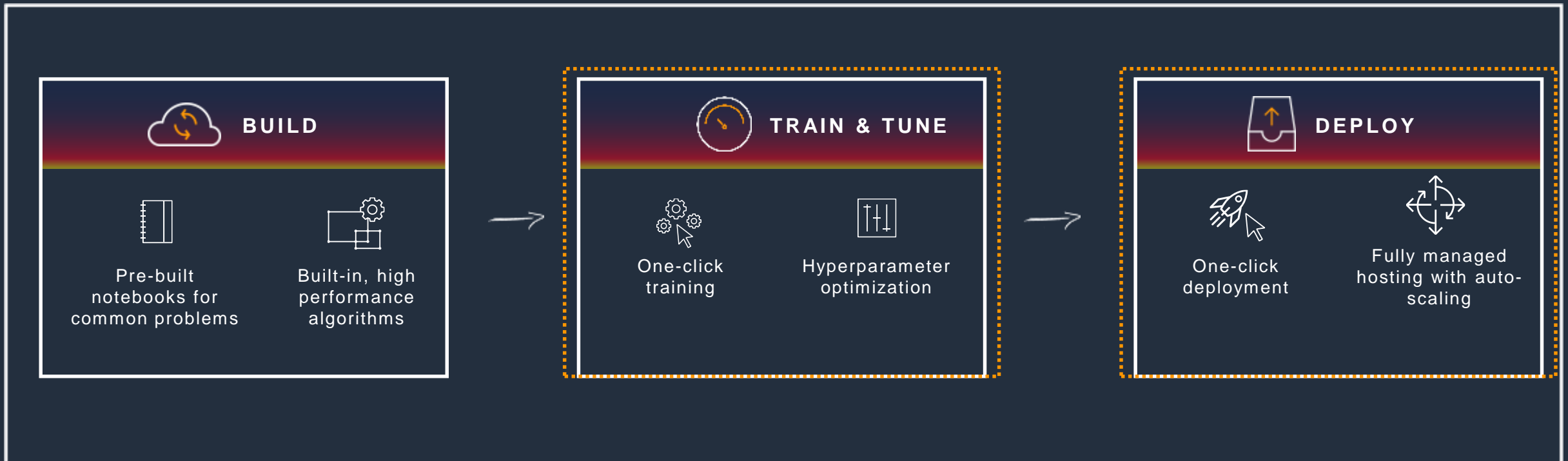
Train

- A powerful instance
- Hyperparameter optimization

Host

- A moderate instance
- Auto-scaling

Amazon SageMaker



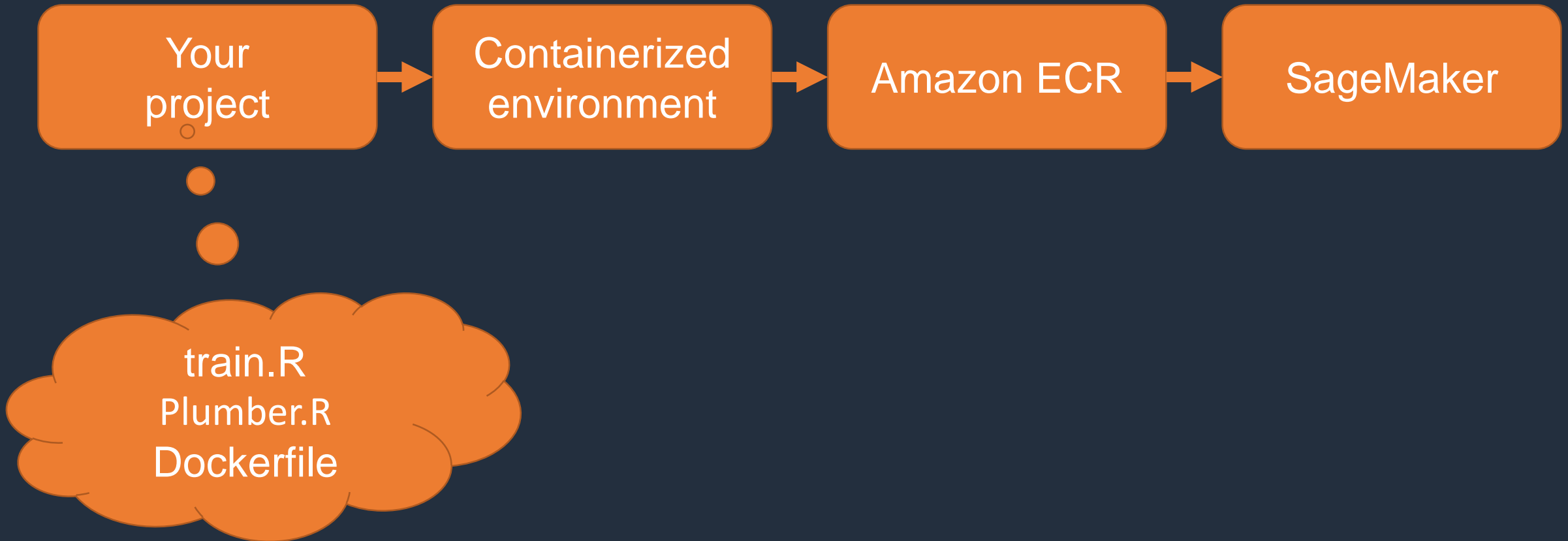
Deploying R models on SageMaker

- ```
train <- function() {

 save(your_model, file='your_model.RData')
}
```
- Run a lightweight HTTP server on a container image for processing requests in hosting (R “Plumber” package).

Source: [https://github.com/aws-labs/amazon-sagemaker-examples/blob/master/advanced\\_functionality/r\\_bring\\_your\\_own/r\\_bring\\_your\\_own.ipynb](https://github.com/aws-labs/amazon-sagemaker-examples/blob/master/advanced_functionality/r_bring_your_own/r_bring_your_own.ipynb)

# How it works?





# EMEA customer Success With Amazon SageMaker



# A take away from this presentation

- Try to containerize your solution (Docker, LXC)
- Try to deploy your model on Amazon SageMaker



Thank you

**EARL**  
CONFERENCE