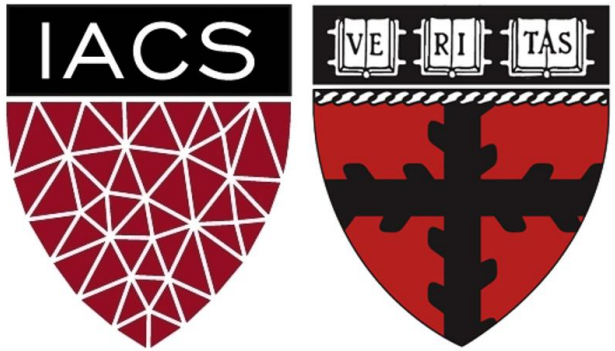
Lecture #1: Introduction

AC295

Advanced Practical Data Science

Pavlos Protopapas



Outline

1 : Why you should take this class (and why not)

2: Who

3: Course Structure and Activities

4: Expectations

5: Workload

6: Logistics

7: Grades

1 : Why you should take this class and (why not)

Learn how to:

put your model in production

integrate and orchestrate applications

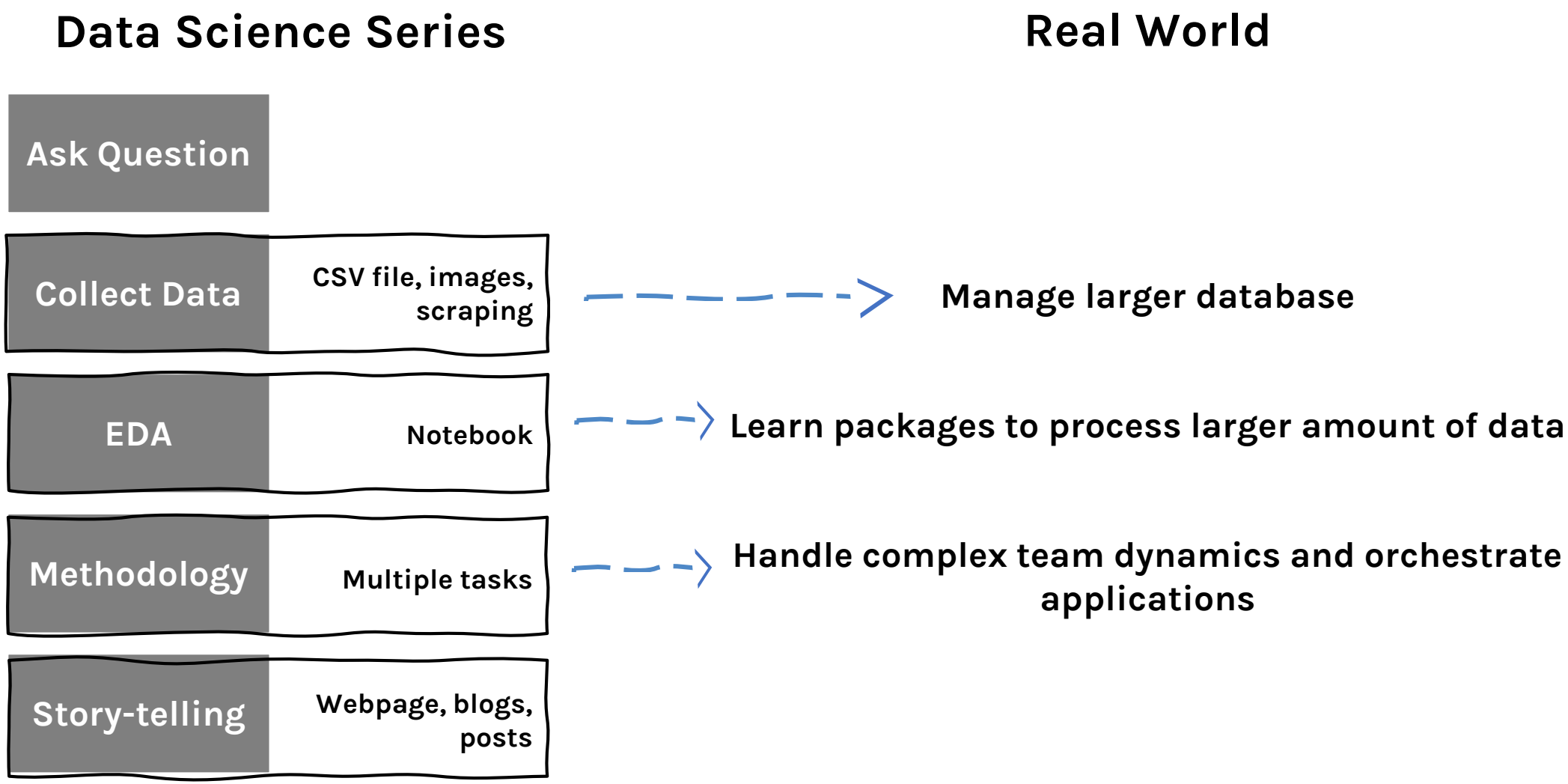
deploy increasing amount of data

take advantage of available models

evaluate and debug model using visualization

> *Syllabus* <

UNK: from Data Science Series to Real World

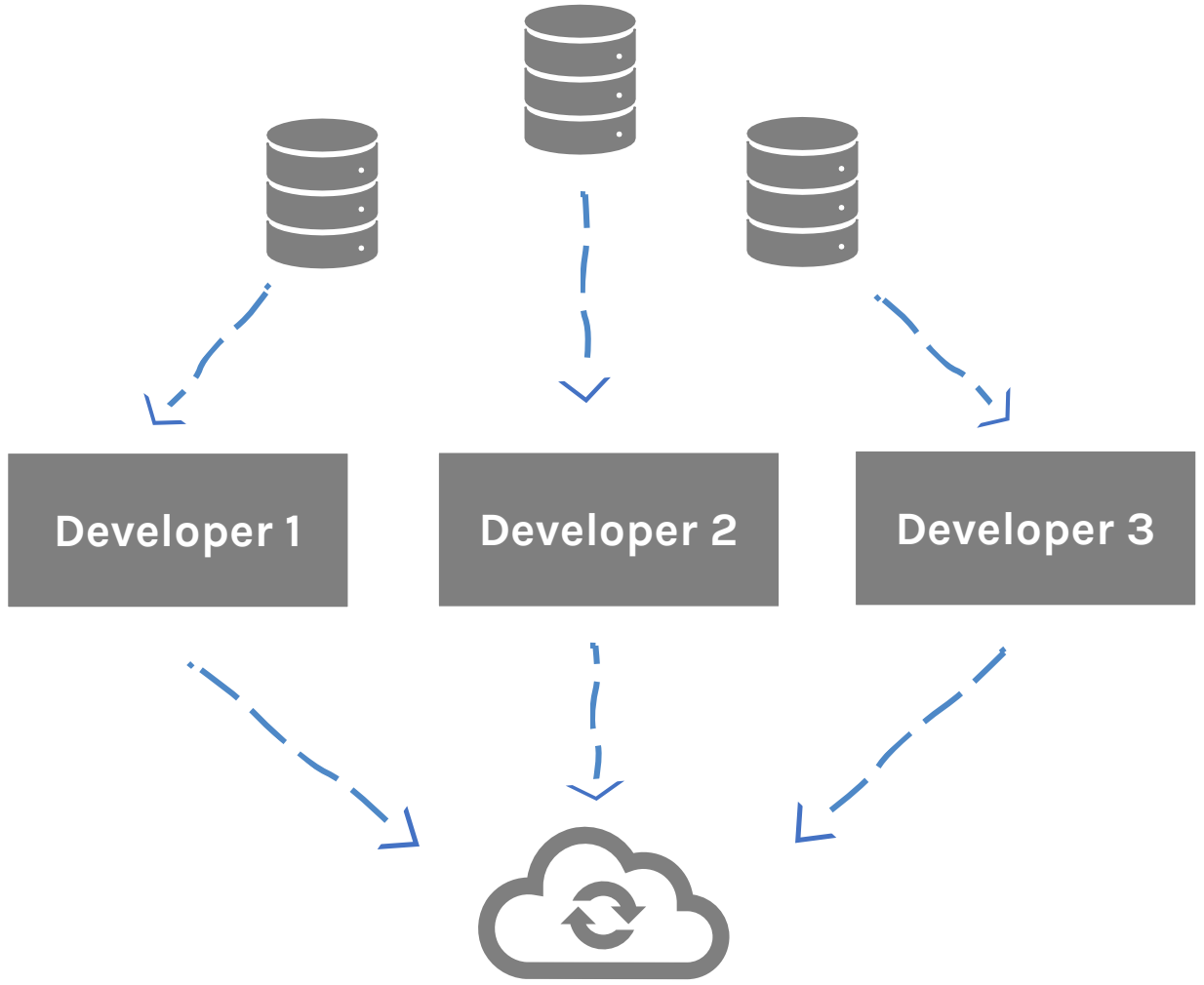


UNK: Real World Example

Fragmented database

Multitude requirements and applications

Recombine and deploy

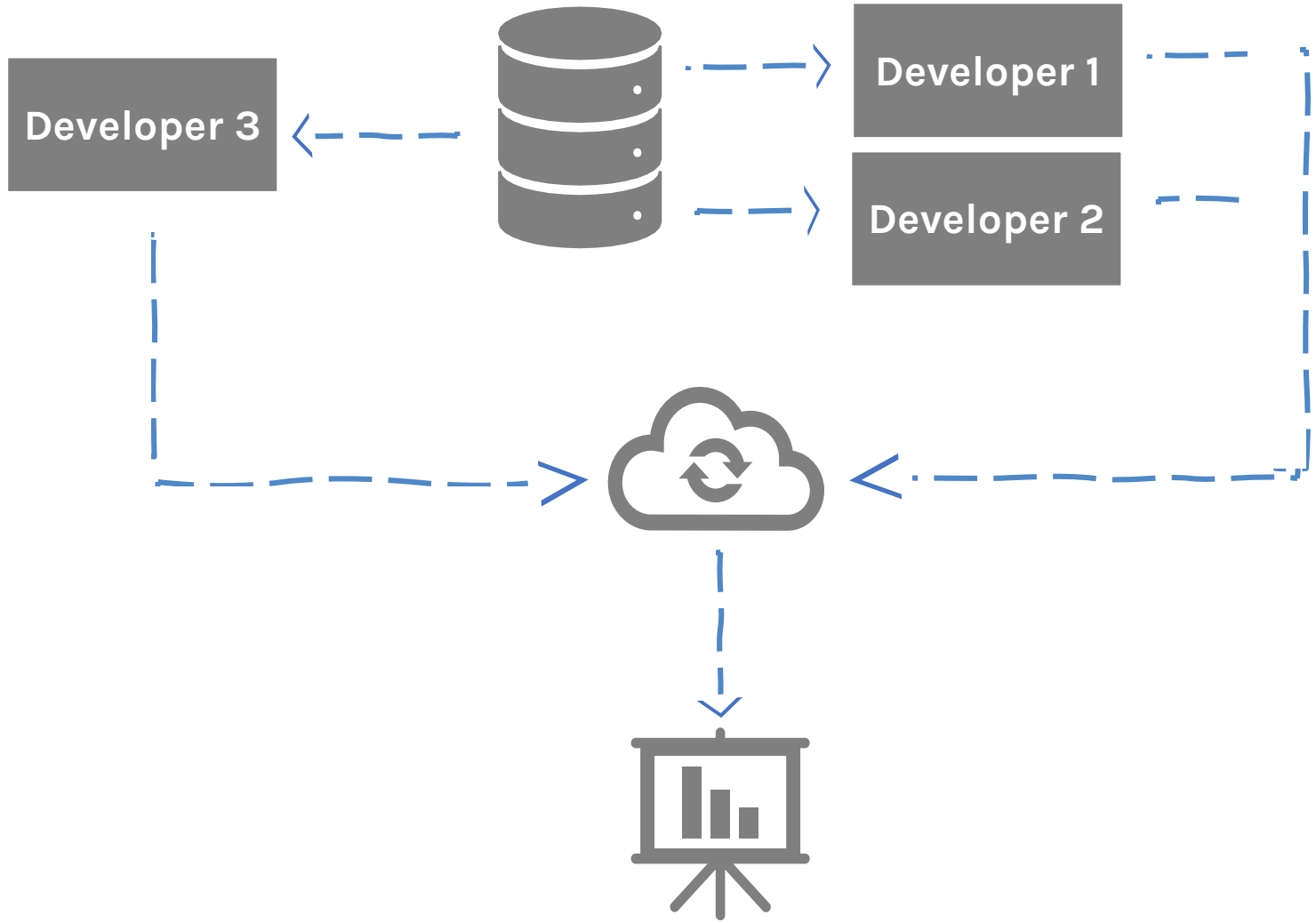


UNK: Real World Example EDA

Huge database or task (i.e. Ensemble)

Recombine results

Present results



2: Who

Pavlos Protopapas

Scientific Director of the Institute for Applied Computational Science (IACS)
Teaches CS109(a/b) and the Capstone course. Research in astrostatistics and he is excited about the new telescopes coming online in the next few years. He has absolutely no hobbies or interests except teaching CS109 and eating.

2: Who

Michael S. Emanuel

After 17 years in finance, mainly fixed income portfolio management, Michael started a second career and is completing the Masters of Data Science program at Harvard. He is a father of two small children who occasionally crash IACS events and enjoys distance running and classical music.

2: Who

Andrea Porelli

Urban planner turned into data hacker. He likes to break things just for the sake of putting them back together (most of the time). Committed to apply Data Science to change something. So far, he managed to change himself the most –thanks IACS- and look forward to pass it over.

2: Who

Giulia Zerbini

Data Designer. Creative technologist at The Visual Agency in Milan, MA Graduate at Politecnico di Milano. Designing and developing visualizations and interfaces based on data. Passionate about using visualizations for discovering patterns in data and communicating information in intuitive terms to a broad audience.

3: Course Structure and Activities

Modules: 1. Deploy data science (integration + scalability)

2. Transfer learning and distillation

3. Visualization as investigative tool

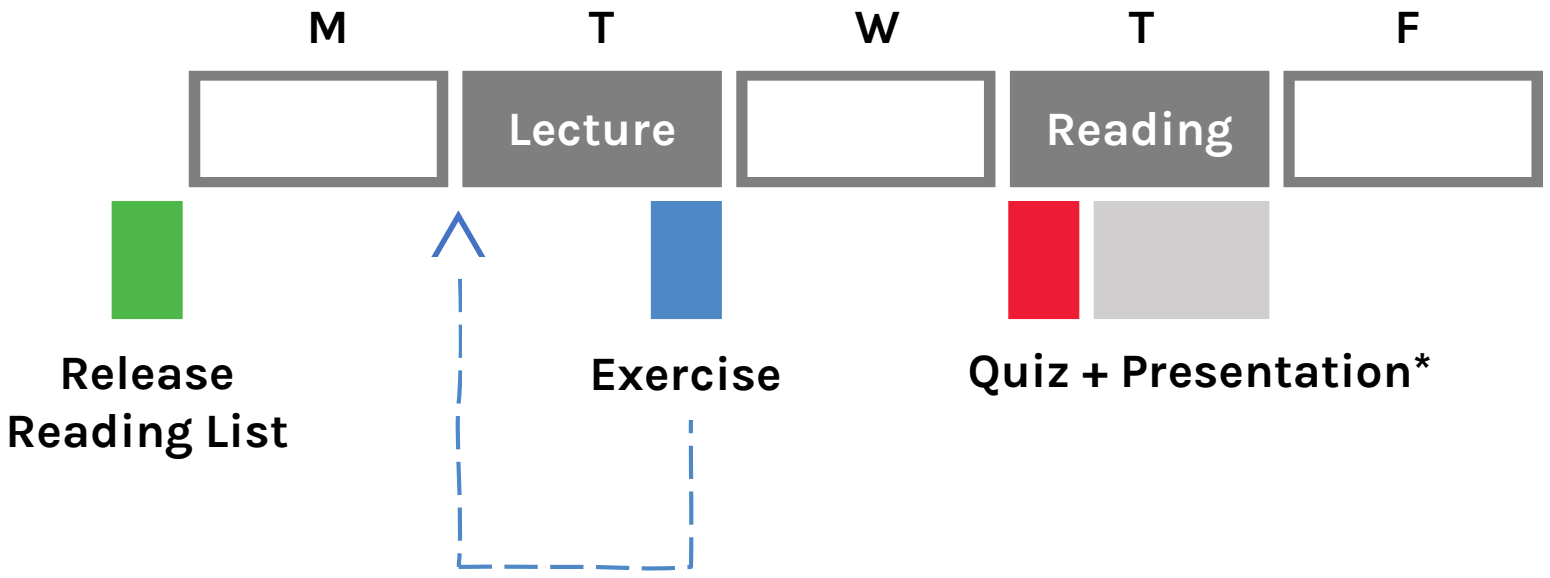
Activities: lectures, reading discussions, practicums, projects

Lectures: Tuesday and Thursday 4:30-5:45 pm in Cruft 309

> *Calendar* <

3: Course Structure and Activities

Regular week schedule



4: Expectations

How to read and present class material

> *Link to Guidelines* <

5: Workload

Regular Week

3 hours in class
3 hours reading
2 hours exercise
2 hours presentation*

~ 10 hours/week

* 1 presentation per module per group (3 total)

Practicum and Project Week

~ 15 hours/week**

* 3 practicums and 1 final project (2 weeks long)

6: Logistics

Fill up forms

Make group *

Sign-up presentation**

* Fill group components in each row

** Each group should pick one slot in each module

7: Grades

Assignment	Final Grade Weight
Quizzes	9%
Exercises	9%
Presentations	15%
Practicums	45%
Projects	20%
Participation	2%
Total	**100%**

AC295

Advanced Practical Data Science

Pavlos Protopapas