Successful Systems in Production (SSP) Graduate Teaching

Ali Shoker
INESC TEC, Portugal

Denver, Nov, 2019

MAPi DOCTORAL PROGRAMME IN COMPUTER SCIENCE

INESCTEC TECHNOLOGY & SCIENCE

Roadmap

- Motivation & Objectives

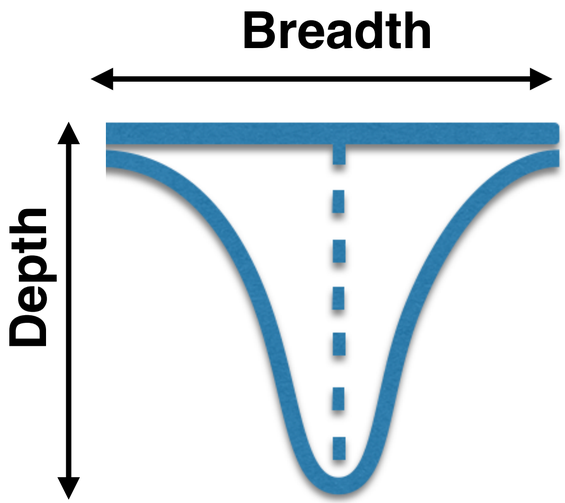
- Course description
    - Two sample topics
    - Methodology
    - Assessment methods

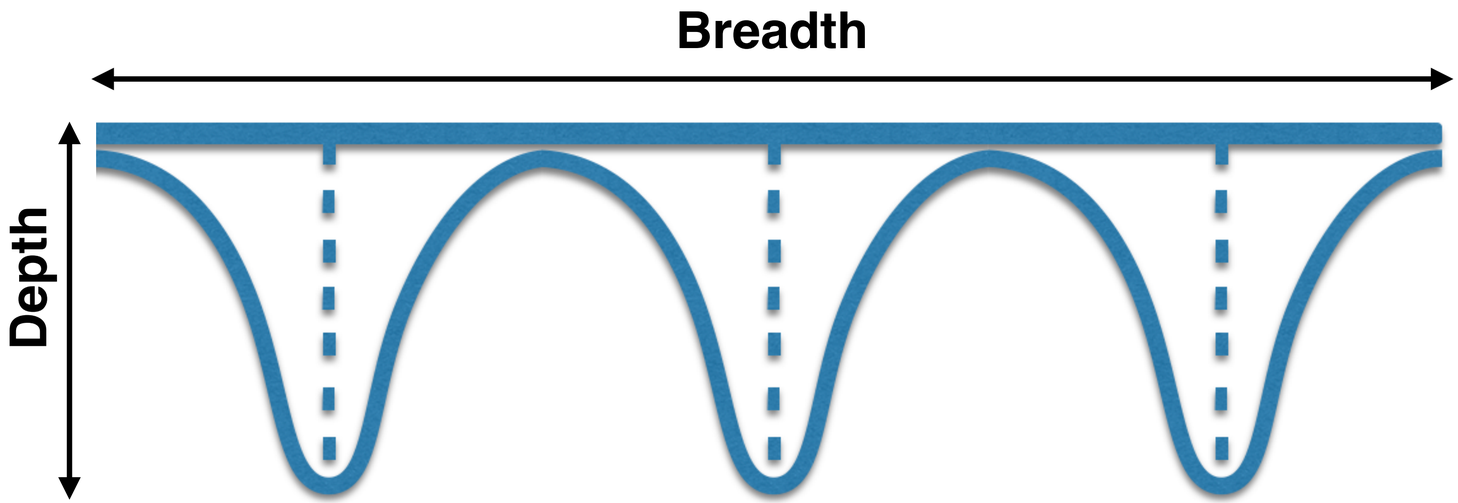
- Evaluation & Challenges

Motivations and objectives

Modern PhD programs

Know everything about something and know something about everything





Theory & Practice: hand in hand

Computing industry is ahead, e.g., **Blockchain**

Focus on **things that work in production**

Objectives are to help graduates:

- overview various cutting edge (Distributed) Systems **in Production**

- inspire from **research & practice**

- and figure out **synergies with own thesis/research**

Course description

- Topics
- Methodology
- Assessment

Successful Systems in Production, are often distributed

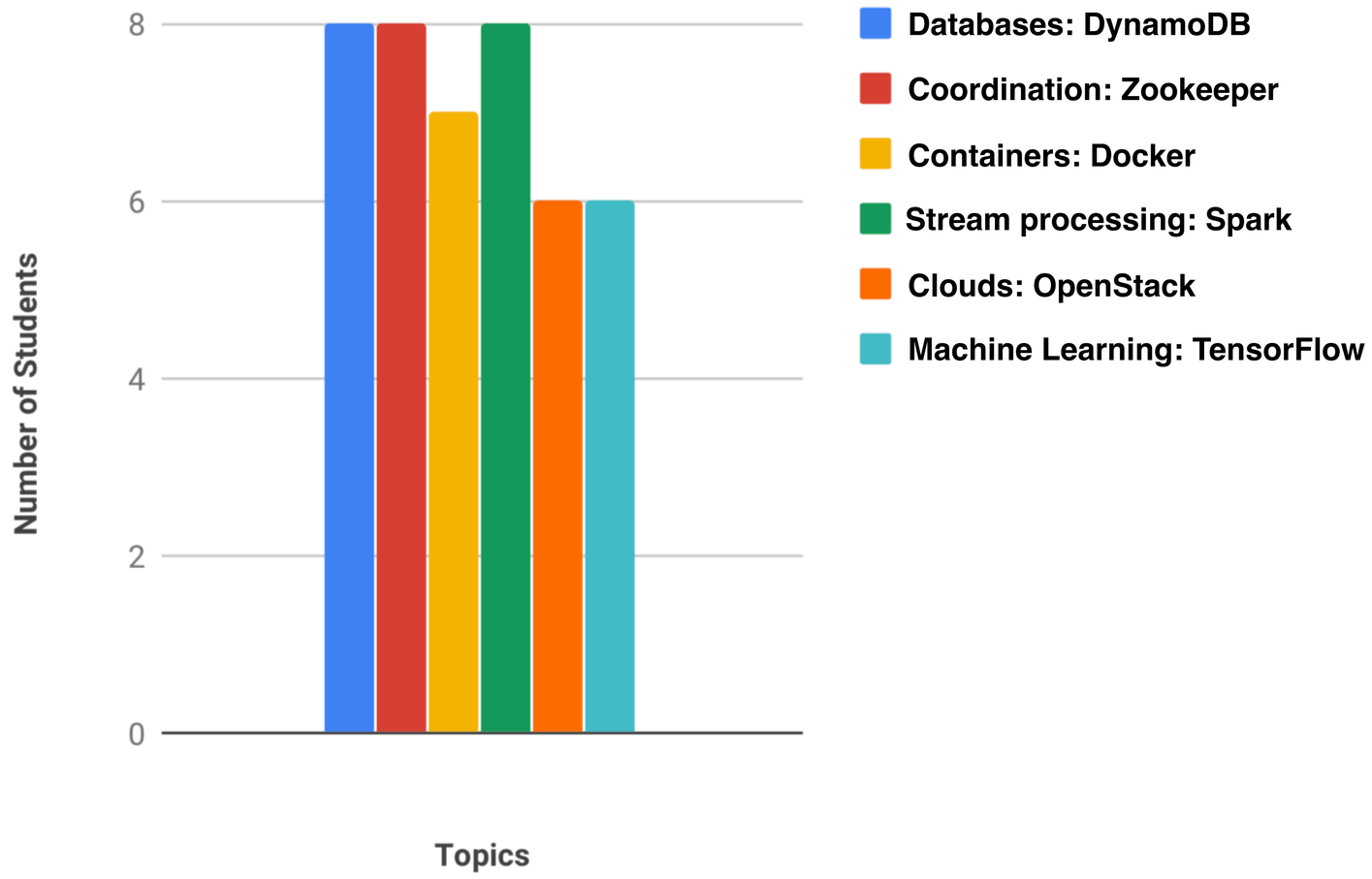
Proposed topics, pick 6

diverse system stack layers

diverse computing areas

Topic	SSP	Area
Databases	Amazon DynamoDB	All
In-Memory Cache	Redis	Sys,DB,BC,CC,SE
Coordination systems	Apache Zookeeper	Sys,BC,SE,CC,Com,Sec
Publish/Subscribe	Twitter	Sys,SE,BC,Com,Sec
Message brokers	Kafka	Sys,AI,CC,Com,SE,Sec
Blockchains	Bitcoin	All
Cloud Computing	OpenStack	Sys,CC,DB,Com,SE,Sec
Containers	Docker	All
Big Data	Hadoop	All
Stream Processing	Spark	Sys,AI,IoT,Sec,SE,Com
Deep Learning	TensorFlow	AI,Sys,Sec,IoT,SE

Topics that “should remain next year”



Teaching methodology emphasis

- **Intersection** with students' **own research**

- Touch upon **major theoretical** works

- Dig in the **practical** aspects of a successful system

  - two models (see next)

Two sample topics

Coordination/Zookeeper

Coordination (theory)
- why distributed?
- clocks
- 2PC
- 3PC
- Paxos

Zookeeper (practice-anatomy)
- architecture
- data model
- ordering
- ZAB protocol
- API
- patterns, notifications, etc.

Containers/Docker

Containers (theory)
- cloud computing
- virtualisation types
- containers
- *cgroups*
- *namespaces*
- containers vs VMs

Docker (practice-hands-on)
- Run an image
- *Dockerfile*
- Interact
- Network
- Distributed

Assessment—inline with thesis

- A project where one SSP topic **intersects** with student's thesis/research

- Main evaluation criteria: how the project **serves the thesis**?

Evaluation & feedback

- Tangible results
- Student’s feedback

Tangible results

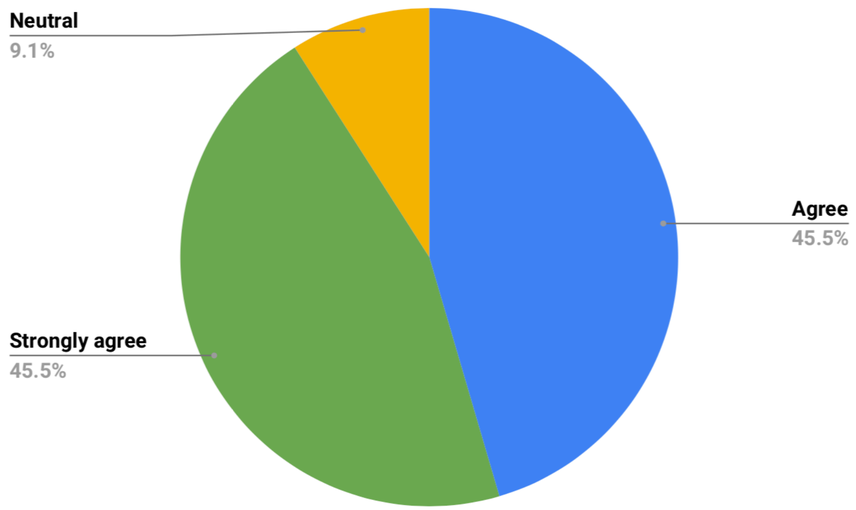
- **Security**: *SafeDynamo*: a proof-of-concept prototype modular security for Dynamo-like databases (Macedo, IEEE SRDS, 2017).

- **Distributed Systems**: *Tricks*: a benchmarking tool for deploying and experimenting distributed computing protocols on *Kubernetes*.
GitHub, 2017: https://github.com/vitorenesduarte/tricks

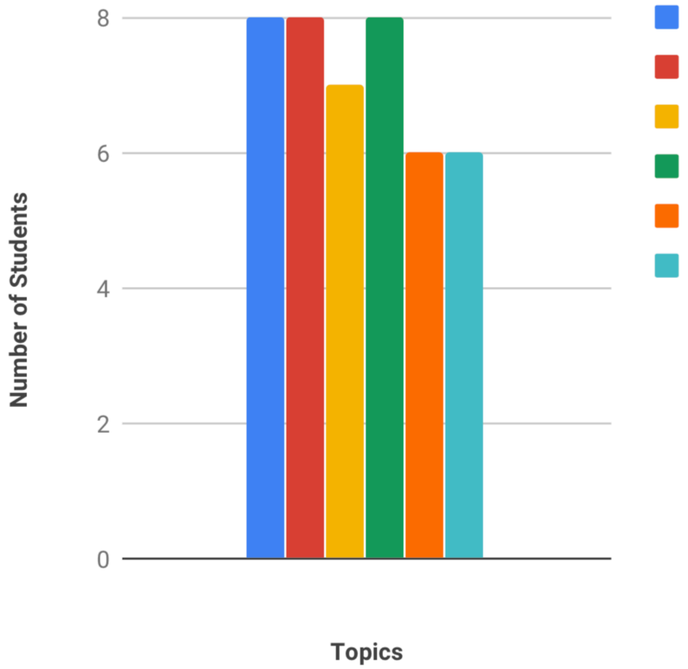
- **Processing**: accuracy and overhead of running *Spark* or *Hadoop* ecosystems in a cluster versus distributed container-based deployments.

What students say?

Most useful course I took in the PhD program?



Topics should remain next year?



Challenges & recommendation

Challenges

- **Centralisation:** may be overwhelming for the coordinator

- **Consistency:** try to impose style/methodology on lectures ;)

- **Preparation:** topics diversity vs time

Highly recommended for

- affordable PhD programs,
- diverse backgrounds, and
- practical in production systems

Happy to help :)

Ali Shoker
shokerali@gmail.com
www.alishoker.com