Successful Systems in Production (SSP)
Graduate Teaching

Ali Shoker
INESC TEC, Portugal

Denver, Nov, 2019
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

<table>
<thead>
<tr>
<th>Topic</th>
<th>SSP</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Databases</td>
<td>Amazon DynamoDB</td>
<td>All</td>
</tr>
<tr>
<td>In-Memory Cache</td>
<td>Redis</td>
<td>Sys, DB, BC, CC, SE</td>
</tr>
<tr>
<td>Coordination systems</td>
<td>Apache Zookeeper</td>
<td>Sys, BC, SE, CC, Com, Sec</td>
</tr>
<tr>
<td>Publish/Subscribe</td>
<td>Twitter</td>
<td>Sys, SE, BC, Com, Sec</td>
</tr>
<tr>
<td>Message brokers</td>
<td>Kafka</td>
<td>Sys, AI, CC, Com, SE, Sec</td>
</tr>
<tr>
<td>Blockchains</td>
<td>Bitcoin</td>
<td>All</td>
</tr>
<tr>
<td>Cloud Computing</td>
<td>OpenStack</td>
<td>Sys, CC, DB, Com, SE, Sec</td>
</tr>
<tr>
<td>Containers</td>
<td>Docker</td>
<td>All</td>
</tr>
<tr>
<td>Big Data</td>
<td>Hadoop</td>
<td>All</td>
</tr>
<tr>
<td>Stream Processing</td>
<td>Spark</td>
<td>Sys, AI, IoT, Sec, SE, Com</td>
</tr>
<tr>
<td>Deep Learning</td>
<td>TensorFlow</td>
<td>AI, Sys, Sec, IoT, SE</td>
</tr>
</tbody>
</table>
Topics that “should remain next year”

- Databases: DynamoDB
- Coordination: Zookeeper
- Containers: Docker
- Stream processing: Spark
- Clouds: OpenStack
- Machine Learning: TensorFlow
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


- **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?

- Agree: 45.5%
- Strongly agree: 42.5%
- Neutral: 9.1%

Topics should remain next year?

[Bar chart showing the number of students for various topics]
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