A Provider-Friendly Serverless Framework for Latency-Critical Applications

Simon Shillaker
Supervisor: Prof. Peter Pietzuch
Group A Planner, Cloud Computing

http://lsds.doc.ic.ac.uk
s.shillaker17@imperial.ac.uk

12th Eurosys Doctoral Workshop, April 23rd, Porto, Portugal
Overview

*Why does anyone care?*

*What's the problem?*

*How will I address this?*

*What are the consequences?*
Why does anyone care?

Serverless makes distributed, highly parallel systems easy
What's the problem?

Variable latency for the user and unpredictable resource requirements for the provider
Serverless Architecture

SCHEDULER

FUNCTION STORE
Function A: {
  Code: ...
  Libraries: ...
}

=  
- Container OS
- Language runtime
- Function code
- Libraries
Overview

**How will I address this?**

Throw out containers and build a new runtime
Serverless Runtime - Architecture
Serverless Runtime – Scheduling

FUNCTION 1
A
B

FUNCTION 2
A
C

SCHEDULER

WORKER 1
RUNTIME
A
B

WORKER 2
RUNTIME
A
D

WORKER 3
Overview

Why does anyone care?
Serverless makes distributed, highly parallel systems easy

What's the problem?
Highly variable latency and unpredictable resource requirements

How will I address this?
Throw out containers and build a new runtime

What are the consequences?
SLOs and latency guarantees with efficient, linear scaling