

OpenDC 2.0

Convenient Modeling and Simulation of Emerging Technologies in Cloud Datacenters



Fabian Mastenbroek

MSc @ TU Delft, AtLarge Research

F.Mastenbroek@atlarge-research.com

CCGrid2021, May 10-13th 2021



and more...



Datacenters Crucial to Today's Society

Cloud industry produces many **widely-used services**

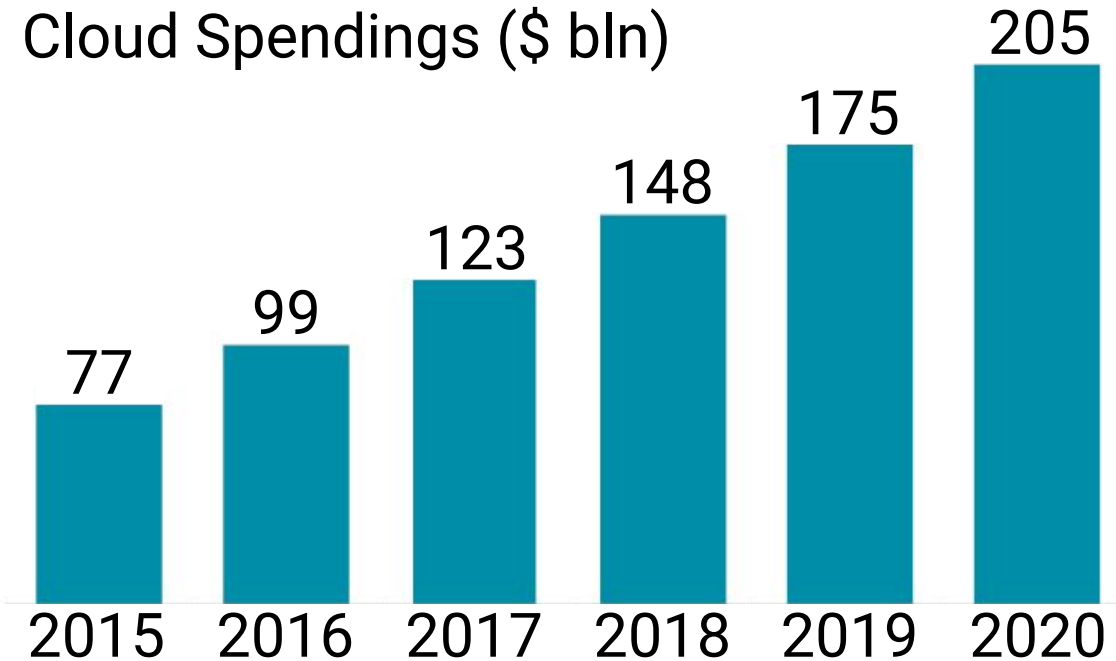


COVID pandemic highlights the importance of digital infrastructure

Gartner Forecasts Worldwide Public Cloud End-User Spending to Grow 23% in 2021

Cloud Spending Driven by Emerging Technologies Becoming Mainstream

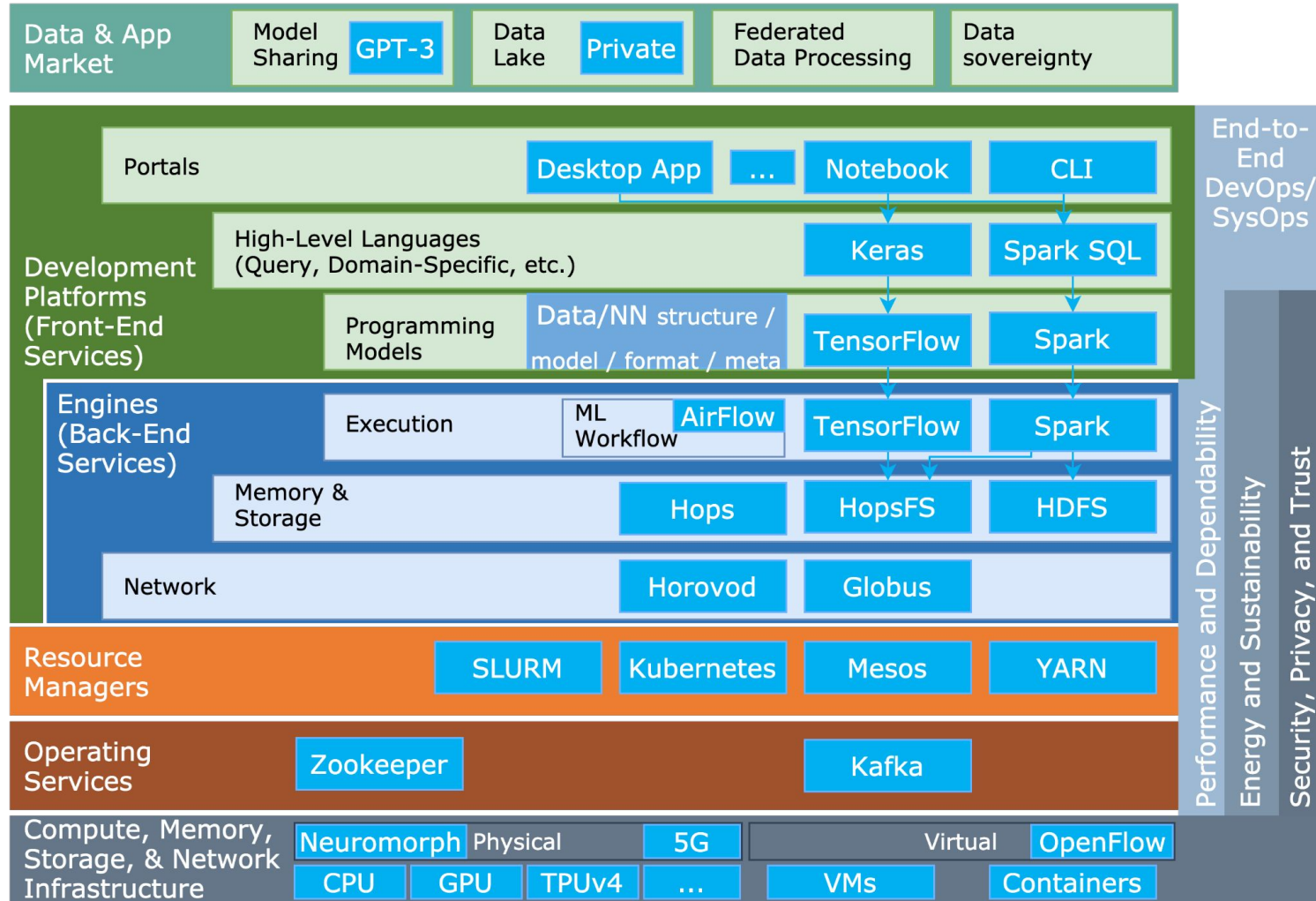
Cloud Spendings (\$ bln)



Source: Nasdaq Global Information Services, 2017 [\[Online\]](#)

Complexity of Datacenters [1/2]

1. Many layers, both hardware & software
2. Many different implementations
3. Many points of failure, issues, bottlenecks



Complexity of Datacenters [2/2]

G Gizmodo Australia

People Can't Vacuum Or Use Their Doorbell Because Amazon's Cloud Servers Are Down

The company that produces Roombas, iRobot, confirmed they (along with their robot mops) were no longer working. An Amazon AWS outage is ...

26 nov. 2020

Data-center outages: Causes are changing, report says

Power issues are less likely to cause a major IT service outage, while IT configuration and network problems are becoming more common, according to the Uptime Institute.

Source: Network World, 2021 [[Online](#)]

The Register
Biting the hand that feeds IT

Google goes dark for 2 minutes, kills 40% of world's net traffic

www.theregister.co.uk/2013/08/17/google_outage/

Systemwide outage knocks every service offline

Source: The Verge, 2020

Source: Guardian, 2020

Source: The Verge, 2020

Sustainability of Datacenters

Power consumption of datacenters:
>1% of global electricity

Source: Nature, 2018 [[Online](#)]

Other greenhouse emissions:
Largely unknown

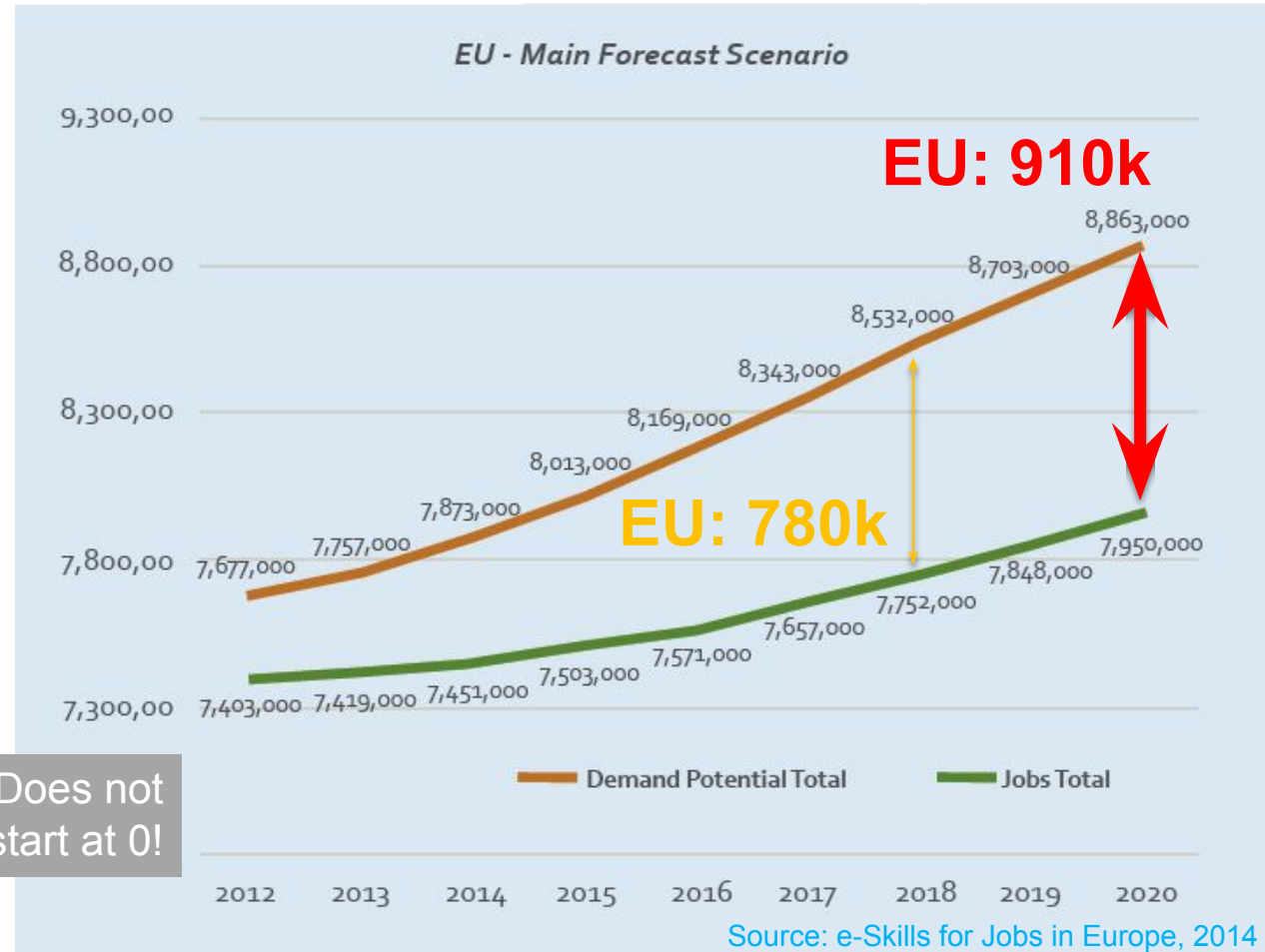
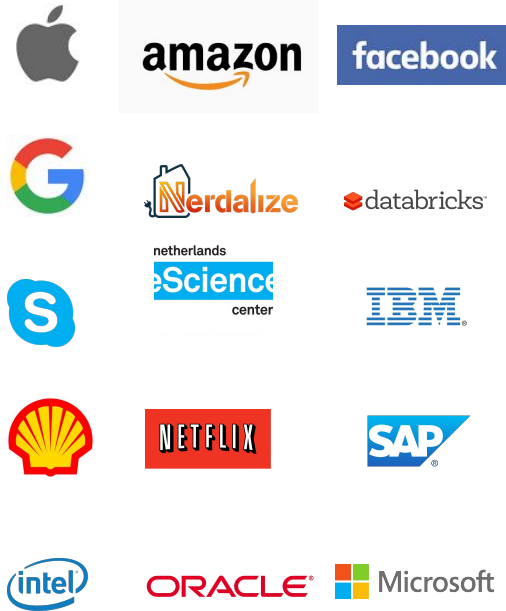
Source: Nature Climate Change, 2020 [[Online](#)]



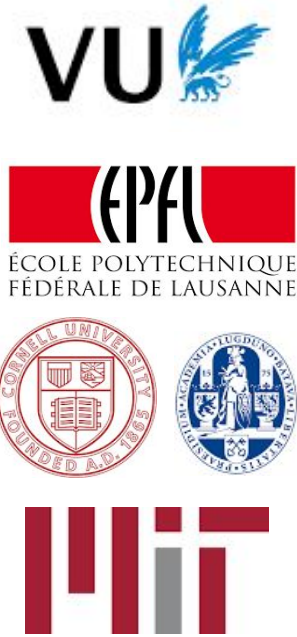
Water consumption of
datacenters
in the US:
>625Bn. l/y (0,1%)

Source: Energy Technologies Area, 2016 [[Online](#)]

Lack of Skilled Personnel



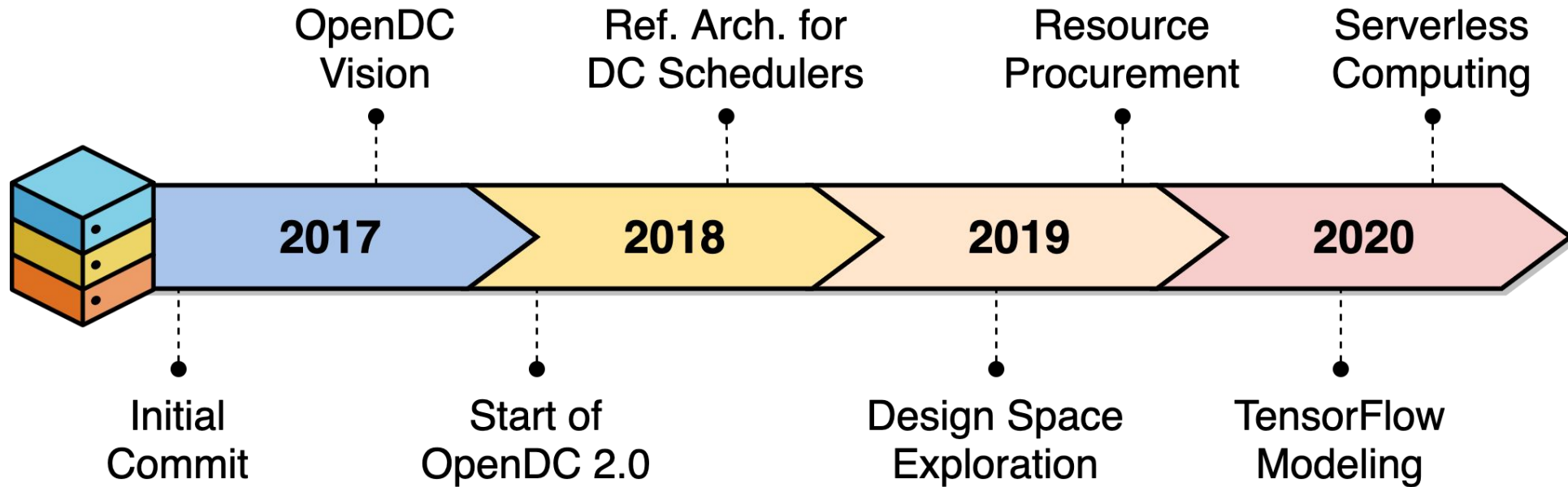
Does not start at 0!



Yet another Simulator?

1. Frequent and substantial innovation in cloud datacenters
2. Difficult to use simulators as platforms for research communities
3. Support different stakeholders

OpenDC 2.0



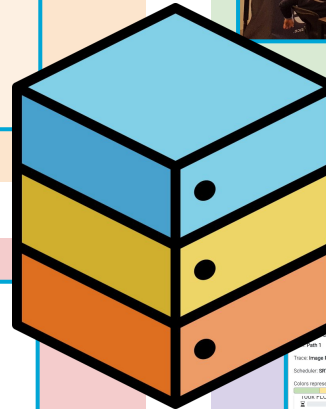
What does OpenDC offer right now?

1. Datacenter Technology & Methods

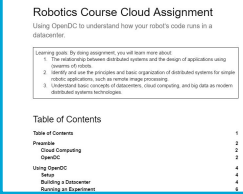
Risk Analysis +
Management

Heterogeneity

Efficiency →
SME
Availability



2. Education Practices



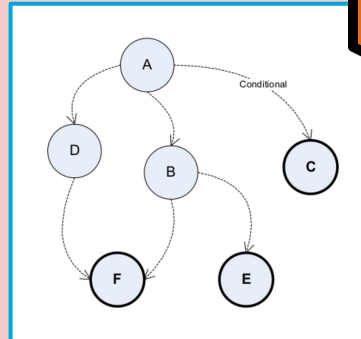
Mnemos: Self-Expressive Management of
Business-Critical Workloads in Virtualized
Datacenters

Vincent van Beek^{1,2} Jesse Donkervliet Tim Hegeman
Stefan Hugtenburg
Alexandru Iosup

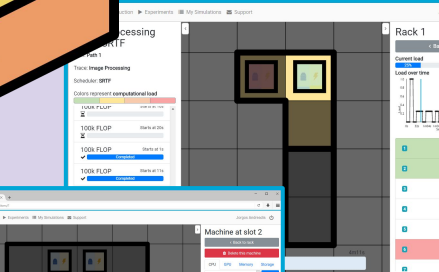
¹ Bitbrains IT Services Inc., Amstelveen, the Netherlands
² Delft University of Technology, Delft, the Netherlands

Corresponding author: vincent.vanbeek@bitbrains.nl

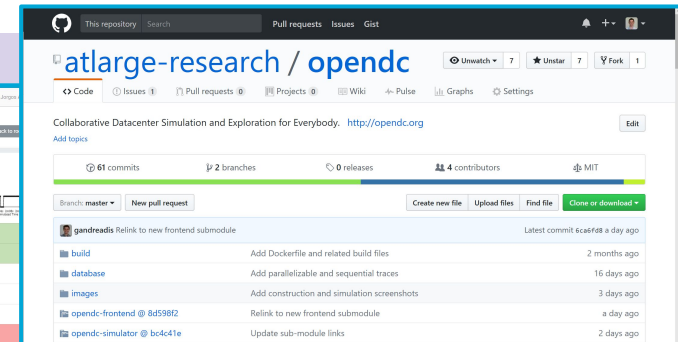
May 18, 2015



3. Scientific Methods



4. Software & Data Artifacts

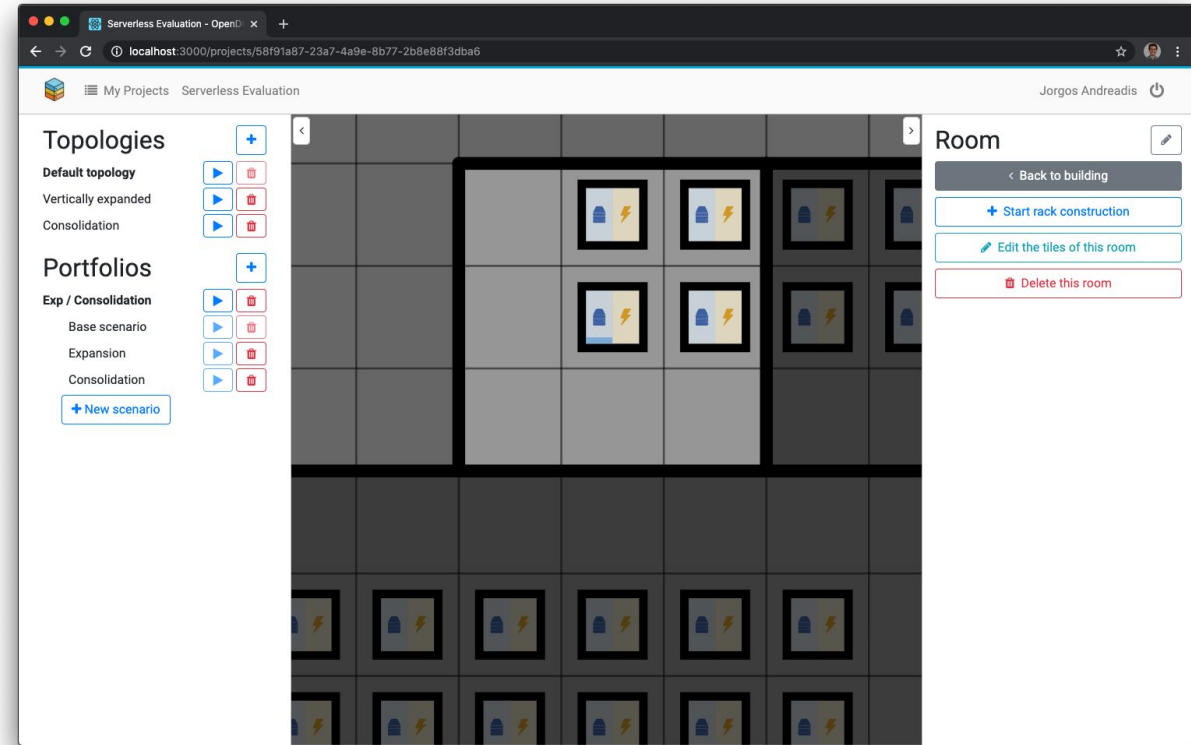


How Does OpenDC work? Foundations of OpenDC

1. Discrete event simulation at its core
2. Generic model for resource-sharing
3. Re-used across components of OpenDC
4. Models operational phenomena in datacenters
 - a. (Un)correlated Failures
 - b. VM Performance Interference
5. Needs to be highly optimized

GUI at the Forefront

1. Design and share datacenters interactively via web interface
2. Conduct online experiments and explore automated plots
3. No user installation necessary

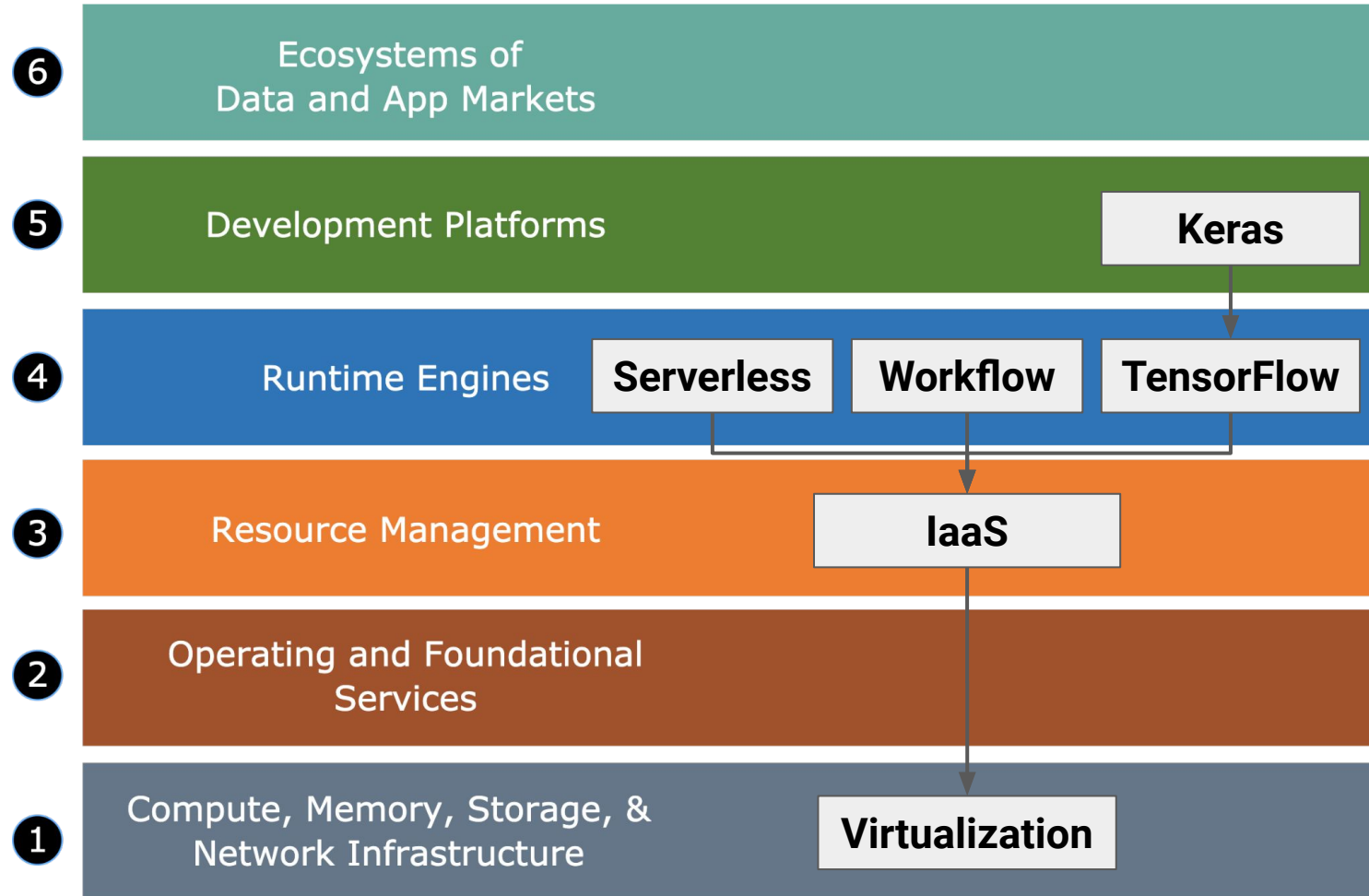


Available as a Service: opendc.org

Capturing the Layers of Datacenters

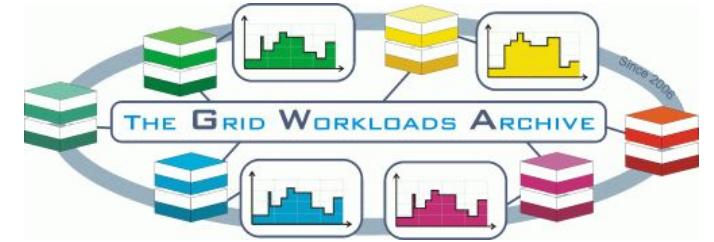
High level models of datacenter operation

1. Resource Mgmt.
2. IaaS
3. Serverless
4. Workflows
5. TensorFlow



Convenience Tools for Simulation

1. Built-in experiment orchestration
2. Support for standardized workload formats
 - a. Grid Workload Archive
 - b. Workflow Trace Archive
 - c. Parallel Trace Archive
3. Extensive set of metrics exposed through industry-initiative OpenTelemetry



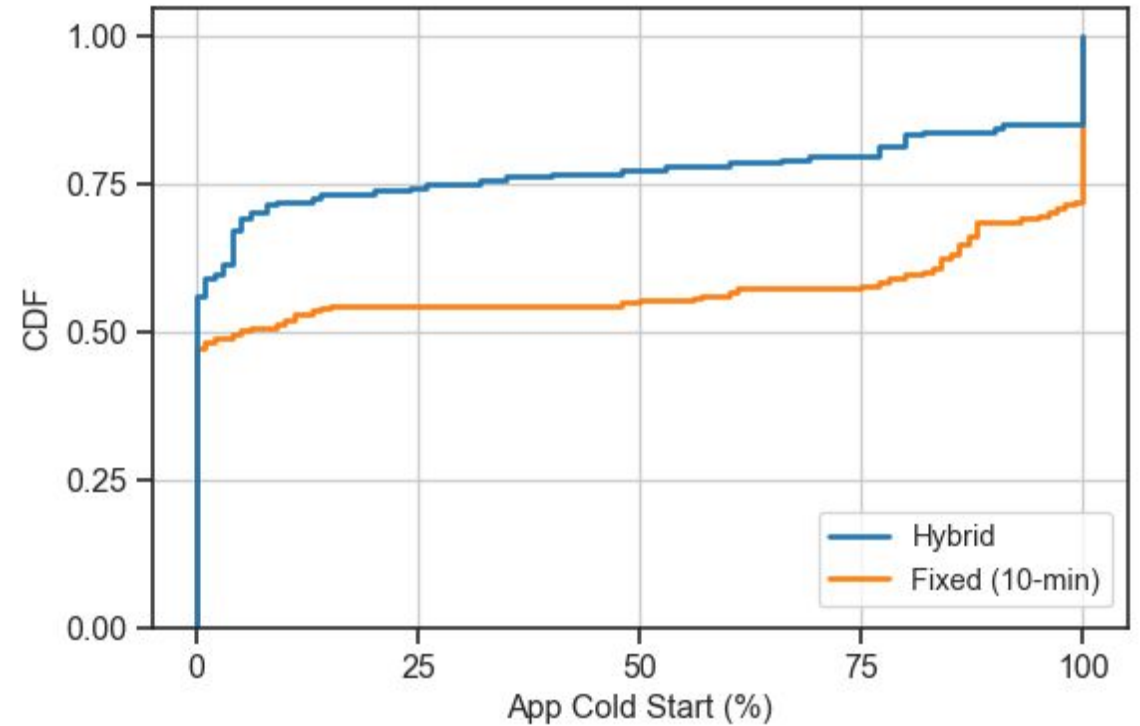
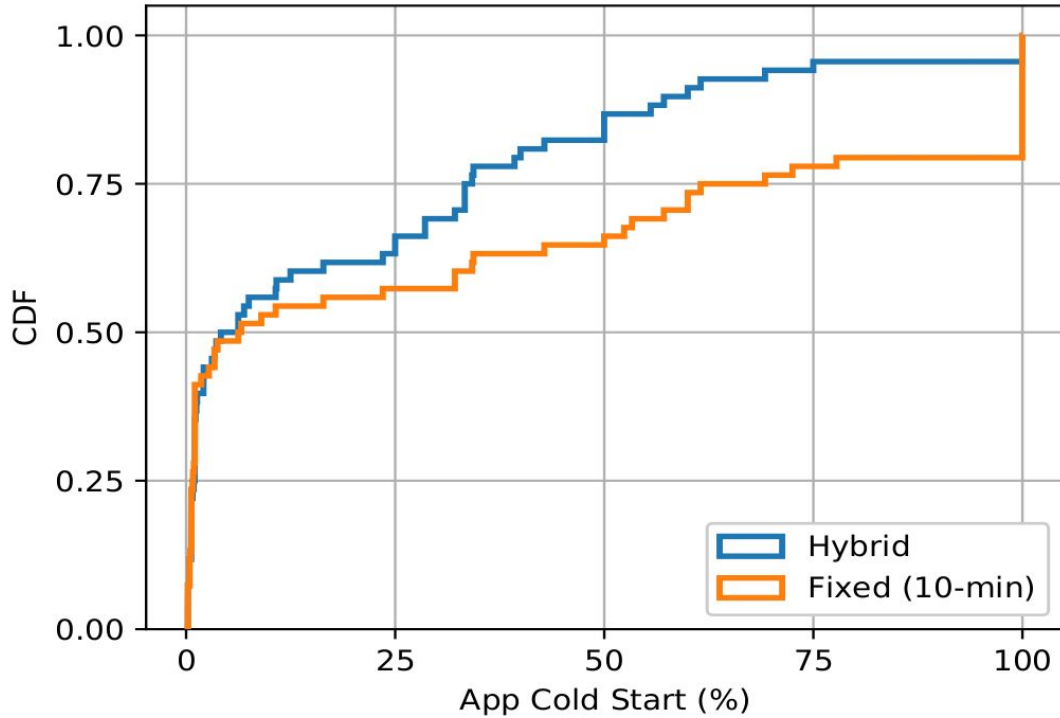
OpenDC 2.0 in Practice: Selected Use-Cases



Full Details:
bit.ly/3xgUeyS

1. Analyzing Serverless Workloads in Simulation
2. Simulation for TensorFlow-based Machine Learning
3. Resource procurement for HPC-as-a-Service with Capelin
4. Reproducibility with and validation of OpenDC
5. Educating Diverse Students on Computer Systems

Analyzing Serverless Workloads in Simulation



Real World

Simulation

Source: Shahradi et al. 2020 [\[Online\]](#)

Does Billing Granularity Matter?

1. AWS Lambda changed billing granularity from 100ms to 1ms in December 2020
2. Cost difference only 1.6% for Azure trace
3. Most functions in this workload have execution times in multiples of a 100
4. Other workloads with small function durations might benefit from this change



Experiments in Simulation: An Environmental Perspective

*Exact numbers confidential,
depend on topology*

1x



Source: Silicon Valley Power, 2021 [\[Online\]](#) [\[Image\]](#)

In Simulation

1x



Source: EIA, 2019 [\[Online\]](#)

In Reality

Take-Home Message

1. Datacenters power today's digital society
2. But we cannot take them for granted
3. OpenDC 2.0 enables convenient simulation of emerging technologies
4. Focus on cloud datacenters
5. Training the next generation of experts



Read our paper:
bit.ly/3xgUeyS



Learn more:
opendc.org



Slides available:
bit.ly/2RZ7msl

Fabian Mastenbroek

F.Mastenbroek@atlarge-research.com

MSc @ TU Delft, AtLarge Research

