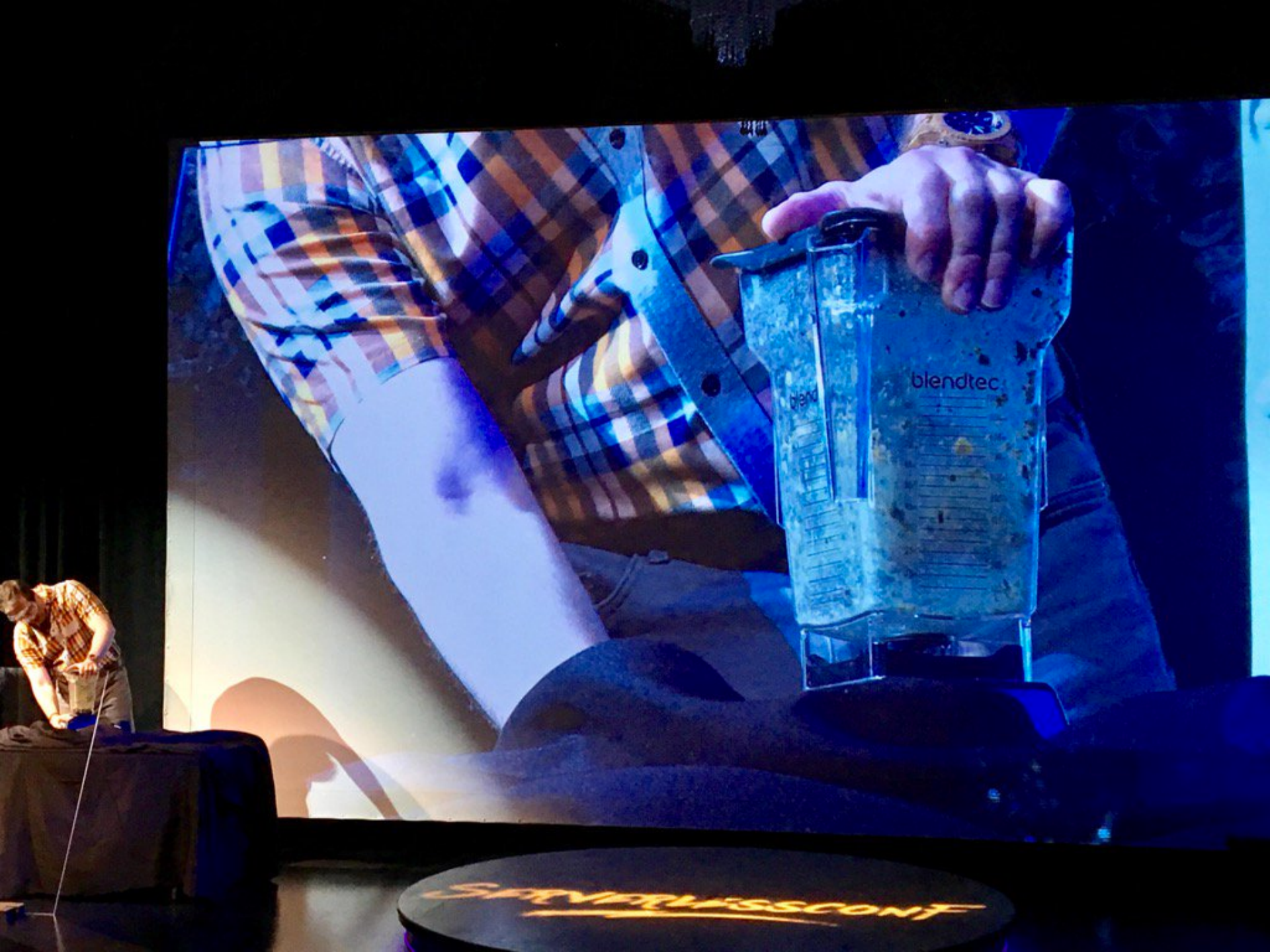
The Serverless Research Opportunity

Johann Schleier-Smith
UC Berkeley

HPTS
October 9, 2017


blendtec

AWS Lambda Selling Points

- *Run code not servers* - no operations needed
- *Elastic scalability* - really big or really small
- *Pay for what you use* - utility billing model

- **Functions** run in response to **events**
- Functions expressed in **popular programming languages**: e.g., JavaScript, Python, Java, C#
- Events include: **web API calls**, **items on queues**, **database triggers**, etc.
- **Cloud provider** provisions runtime environment, charges for each invocation
- Most often used at **“glue”** connecting services

Sample Serverless Uses

- Stateless web API serving
- Image processing, e.g., generating thumbnails or reading checks
- “Cron jobs”
- Sending e-mail / SMS messages
- Event-driven processing pipelines

Google Cloud Platform

OpenLambda

Auth0

Serverless Platform

=

“Serverless” is trending

Source: Google Trends

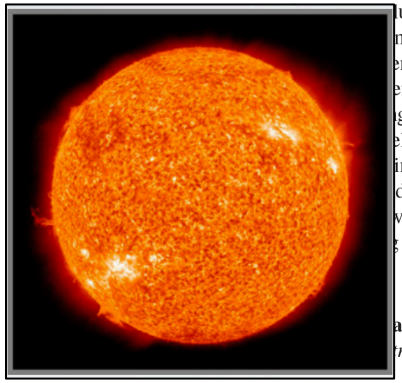
How far can we push serverless?

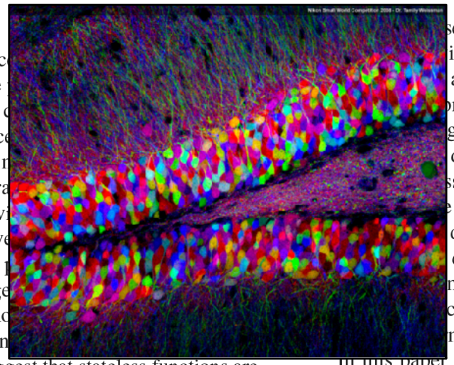
- Big data analytics??
- OLTP????

SoCC 2017 Best Vision Paper

Occupy the Cloud: Distributed Computing for the 99%

Eric Jonas, Qifan Pu, Shivaram Venkataraman, Ion Stoica, Benjamin Recht
University of California, Berkeley
{jonas, qifan, shivaram, istoica, brecht}@eecs.berkeley.edu





ABSTRACT

Distributed computing remains inac
users, in spite of many open source
mercial offerings. While distributed
moved beyond a simple map-reduc
luster
mbarra
nt a v
nt ov
g our
l is g
ing m
ds in
we suggest that stateless functions are
in future computing environments.

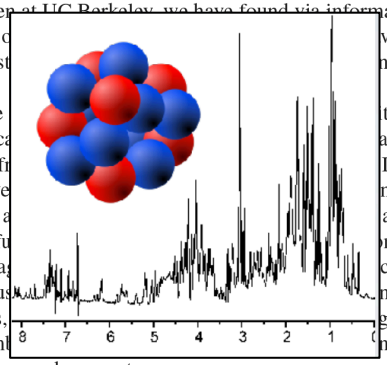
KEYWORDS

Serverless, Distributed Computing, AWS Lambda, PyWren

ACM Reference Format:
Eric Jonas, Qifan Pu, Shivaram Venkataraman, Ion Stoica, Benjamin Recht University of California, Berkeley {jonas, qifan, shivaram, istoica, brecht}@eecs.berkeley.edu 2017 Occupy the Cloud: Distributed Computing

ation → **Cloud computing**; • **Com-**
ributed programming languages;

se installations at large scale. On commercial cloud
ice user confronts a dizzying array of potential deci-
ahead of time decide on instance type, cluster size,
rogramming model, and task granularity.
ges are particularly surprising considering that the
data analytic and scientific computing workloads
ssingly parallel. Hyperparameter tuning for machine
Carlo simulation for computational physics, and
data science all fit well into a traditional map-reduce
even at UC Berkeley, we have found via informal
majo
clust
ms.
In this paper we
stateless functions ca
tic, and more user-fr
this model, we have
that are executed in a
all the state for the fu
shared remote stora
degradation from us
workloads and thus,
to implement a numb
including MapReduce and parameter servers.



Recently cloud providers (e.g., AWS Lambda, Google Cloud Functions) and open source projects (e.g., OpenLambda [16], Open-