Serverless Computing (Lecture 18, cs262a)

Ali Godsi & Ion Stoica,
UC Berkeley
October 28, 2020

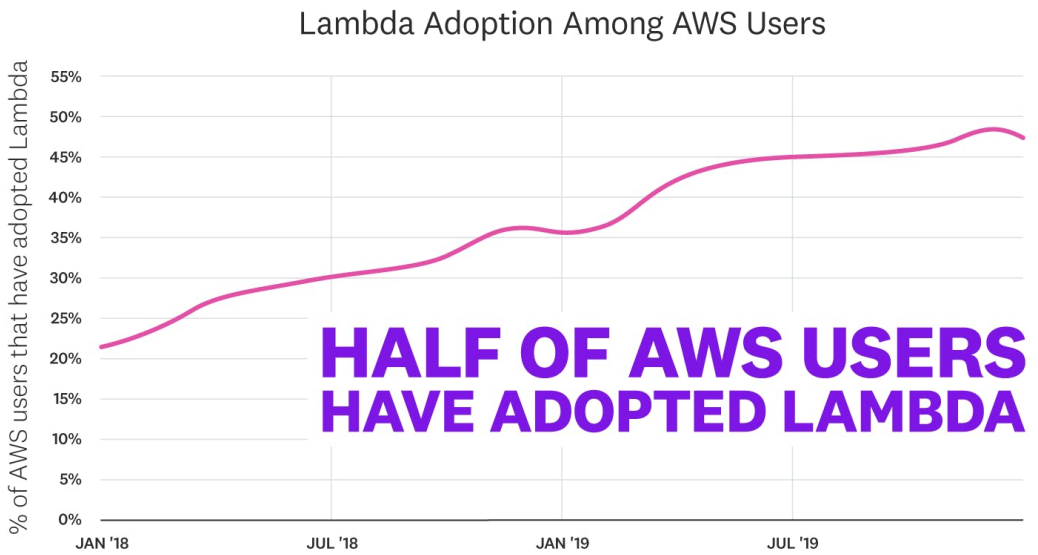
Papers

"Cloud Programming Simplified: A Berkeley View on Serverless Computing", Eric Jonas, Johann Schleier-Smith, Vikram Sreekanti, Chia-Che Tsai, Anurag Khandelwal, Qifan Pu, Vaishaal Shankar, Joao Menezes Carreira, Karl Krauth, Neeraja Yadwadkar, Joseph Gonzalez, Raluca Ada Popa, Ion Stoica and David A. Patterson (https://ucbrise.github.io/cs262a-fall2020/)

"Cloudburst: Stateful Functions-as-a-Service", Vikram Sreekanti, Chenggang Wu, Xiayue Charles Lin, Johann Schleier-Smith, Jose M. Faleiro, Joseph E. Gonzalez, Joseph M. Hellerstein and Alexey Tumanov, (https://arxiv.org/abs/2001.04592)

Why care?

Rapid growth



(from https://www.datadoghq.com/state-of-serverless/) Source: Datadog

Change the way we write applications and expose new challenges

"The future of AWS"

– Marvin Theimer, Distinguished Engineer at AWS

Problem: building distributed apps is hard!

Need to deal with failures

Need to deal with consistency

Need to manager instances:

- What type of instances?
- How many instances?
- What price point?
- Scale up and down # of instances with the demand
- Wait for instances to start…

AWS instance types

EC2Instances.info Easy Amazon EC2 Instance Comparison

Tweet Star

Last Update: 2020-10-28 02:03:47 UTC

351 instances!

EC2 RDS

Region: US East (N. Virginia) | Cost: Hourly | Reserved: 1-year - No Upfront | Columns | Compare Selected | Clear Filters | CSV

Filter: Min Memory (GiB): 0 Min vCPUs: 0 Min Storage (GiB): 0 **Search:**

Name	API Name	Memory	vCPUs	Instance Storage	Network Performance	Linux On Demand cost	Linux Reserve
Search	Search	Search	Search	Search	Search	Search	Search
M5DN Extra Large	m5dn.xlarge	16.0 GiB	4 vCPUs	150 GiB NVMe SSD	Up to 25 Gigabit	$0.272000 hourly	$0.171000 hour
M5A Double Extra Large	m5a.2xlarge	32.0 GiB	8 vCPUs	EBS only	Up to 10 Gigabit	$0.344000 hourly	$0.217000 hour
R5N 12xlarge	r5n.12xlarge	384.0 GiB	48 vCPUs	EBS only	50 Gigabit	$3.576000 hourly	$2.253000 hour
R5AD Extra Large	r5ad.xlarge	32.0 GiB	4 vCPUs	150 GiB NVMe SSD	Up to 10 Gigabit	$0.262000 hourly	$0.165000 hour
R5N Extra Large	r5n.xlarge	32.0 GiB	4 vCPUs	EBS only	Up to 25 Gigabit	$0.298000 hourly	$0.188000 hour
R5DN Extra Large	r5dn.xlarge	32.0 GiB	4 vCPUs	150 GiB NVMe SSD	Up to 25 Gigabit	$0.334000 hourly	$0.210000 hour
I2 Extra Large	i2.xlarge	30.5 GiB	4 vCPUs	800 GiB SSD	Moderate	$0.853000 hourly	$0.424000 hour
M5N 16xlarge	m5n.16xlarge	256.0 GiB	64 vCPUs	EBS only	75 Gigabit	$3.808000 hourly	$2.399000 hour
T2 Micro	t2.micro	1.0 GiB	1 vCPUs for a 2h 24m burst	EBS only	Low to Moderate	$0.011600 hourly	$0.007200 hour
D2 Eight Extra Large	d2.8xlarge	244.0 GiB	36 vCPUs	48000 GiB (24 * 2000 GiB HDD)	10 Gigabit	$5.520000 hourly	$3.216000 hour

https://ec2instances.info/

Serverless

Abstract away servers / clusters:

Pay for computation rather than reserved resources

Two kinds of serverless

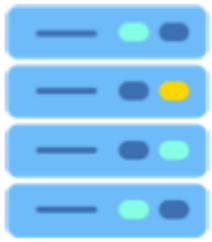
Backend-as-a-Service (BaaS)

- E.g., Big Query, Athena, Databricks
- Users run distributed apps/jobs without reserving machines

Functions-as-a-Service (FaaS)

- E.g., Lambda, Cloud Functions
- Developers build distributed apps without reserving machines

Putting it together



Bare Metal

Code

App Container

Language Runtime

Operating System

Hardware

Virtual machines

Code

App Container

Language Runtime

Operating System

Hardware

Containers

Code

App Container

Language Runtime

Operating System

Hardware

Functions

Code

App Container

Language Runtime

Operating System

Hardware

What does it do?

1. Manage a set of user defined functions
2. Take an event sent over HTTP or received from an event source
3. Determine function(s) to which to dispatch the event
4. Find an existing instance of function or create a new one
5. Send the event to the function instance
6. Wait for a response
7. Gather execution logs
8. Make the response available to the user
9. Stop the function when it is no longer needed