

# MICROSERVICES & TERAFLOPS

## Effortlessly scaling data science with **PyWren**

### #thecloudistoodamnhard



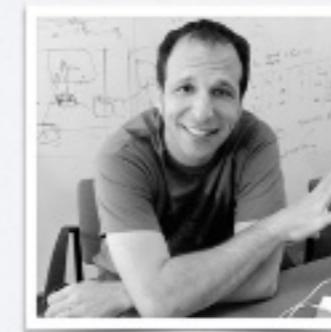
Shivaram Venkataraman  
PhD Candidate, UC Berkeley



Eric  
Jonas



Qifan  
Pu



Ben  
Recht



Ion  
Stoica



Berkeley  
Center for  
Computational  
Imaging

# WHO AM I ?

PhD candidate. AMPLab, RISELab @ UC Berkeley

Dissertation: System design for large scale machine learning

Apache Spark PMC Member

Contributions to Spark core, MLlib, SparkR

# POPULAR SCIENCE

THE  
FUTURE  
NOW

## THE CONTROL CENTERS

Using Data to Feed the World,  
Solve Cold Cases, Battle Malware,  
Predict Our Fate p.52

## OFFICER ALGORITHM

Can a Crime Be Prevented  
Before It Begins? p.38

## NEW WAYS OF SEEING

A Gallery of  
Extraordinary  
Infographics p.68

SPECIAL ISSUE

# DATA IS POWER

HOW INFORMATION  
IS DRIVING  
THE FUTURE

PLUS

Juan Enriquez  
Reprograms Life  
p.31

James Gleick  
Unsplits the Bit  
p.58

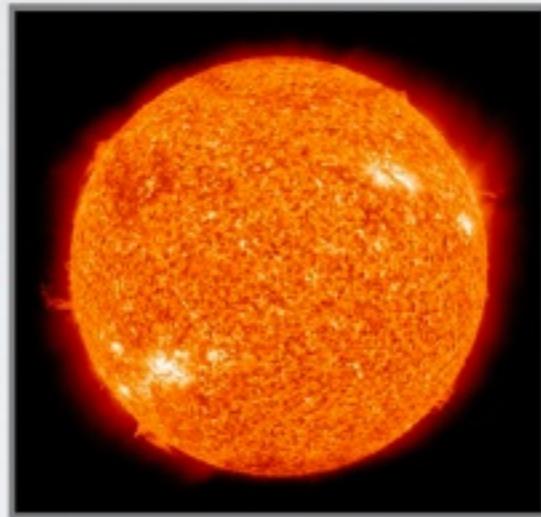
AND  
Lawrence  
Weschler  
Questions the  
Cloud  
p.76



# “BIG” DATA

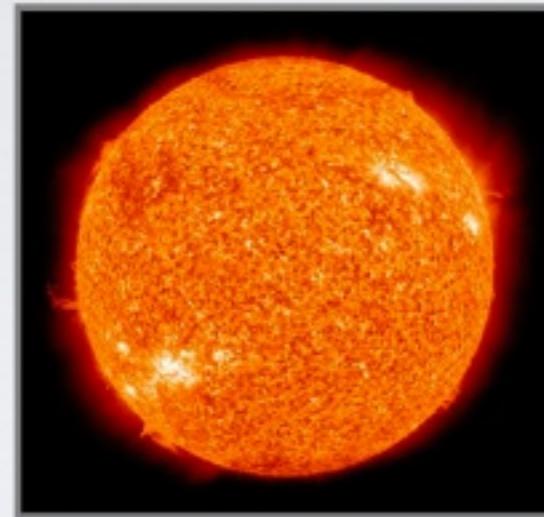
# “BIG” DATA

(near-by) stars

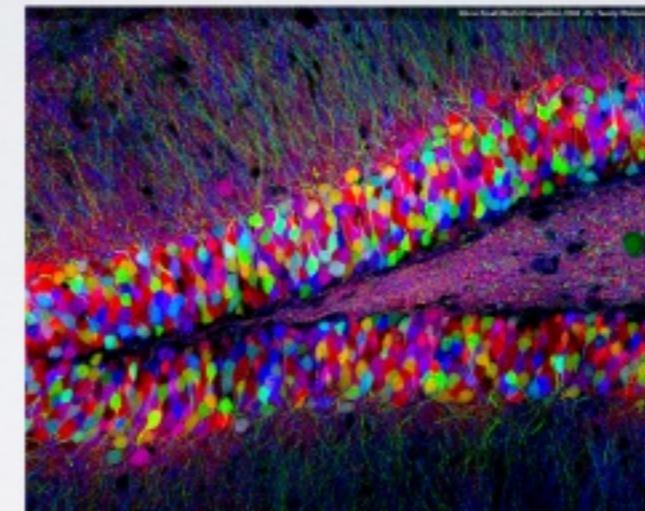


# “BIG” DATA

(near-by) stars

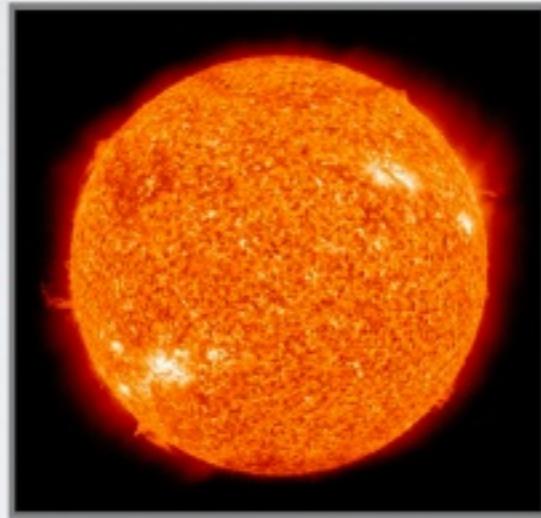


neurons

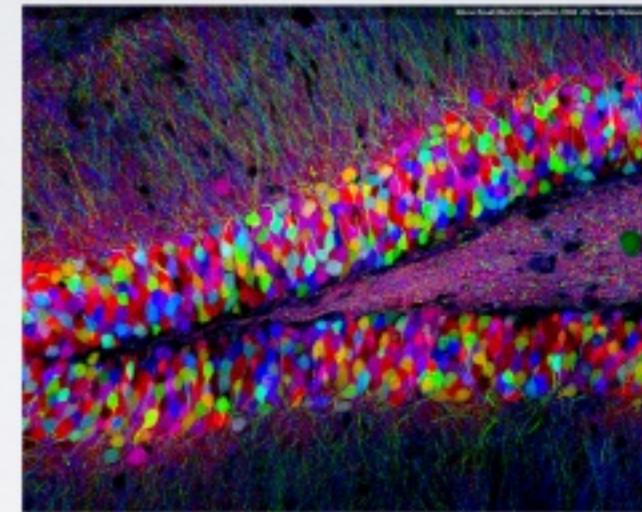


# “BIG” DATA

(near-by) stars



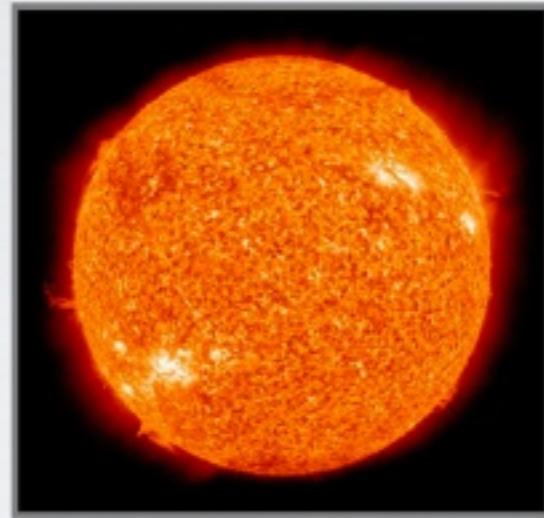
neurons



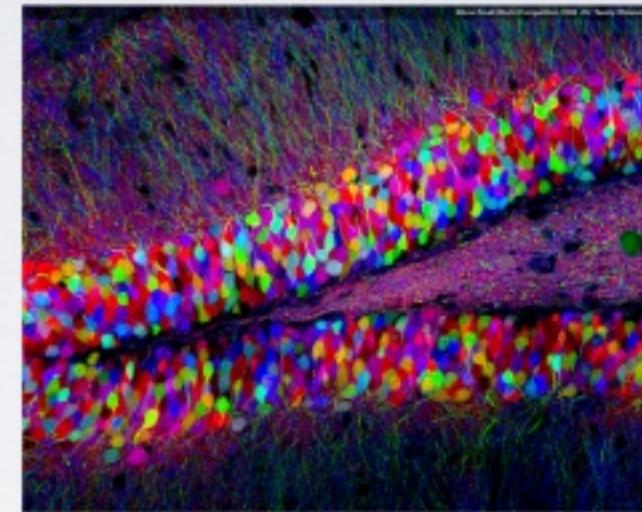
|  |  |  |
|--|--|--|
|  |  |  |
|  |  |  |
|  |  |  |

# “BIG” DATA

(near-by) stars



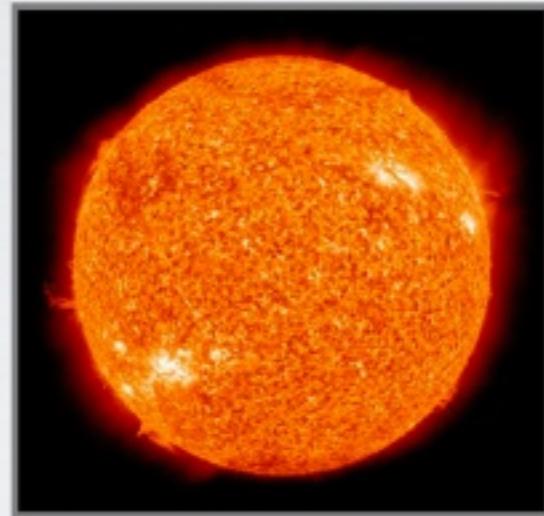
neurons



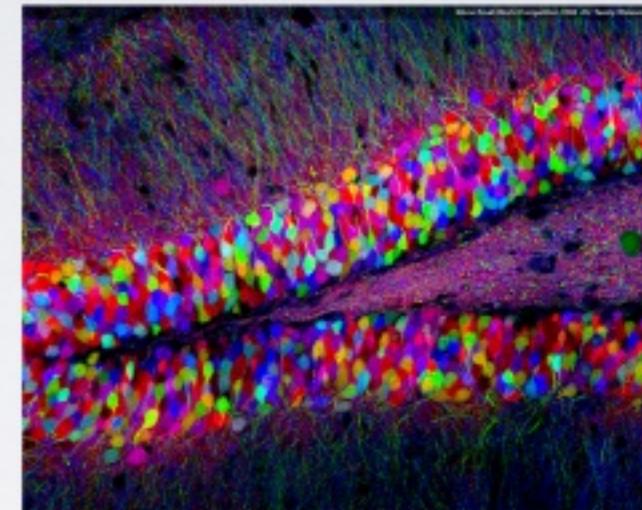
|                  |  |  |
|------------------|--|--|
| <b>size</b>      |  |  |
| <b>number</b>    |  |  |
| <b>data size</b> |  |  |

# “BIG” DATA

(near-by) stars



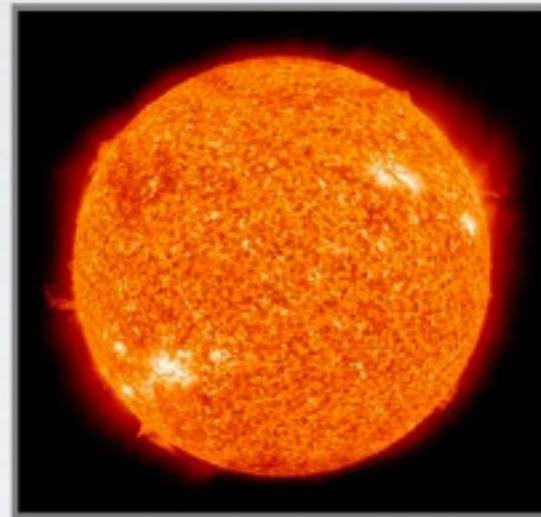
neurons



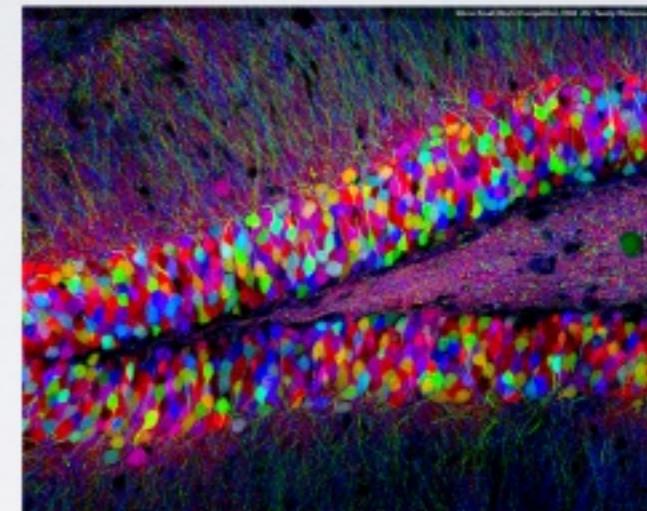
|           |          |
|-----------|----------|
| size      | $10^9$ m |
| number    | 1        |
| data size | 2 PB     |

# “BIG” DATA

(near-by) stars

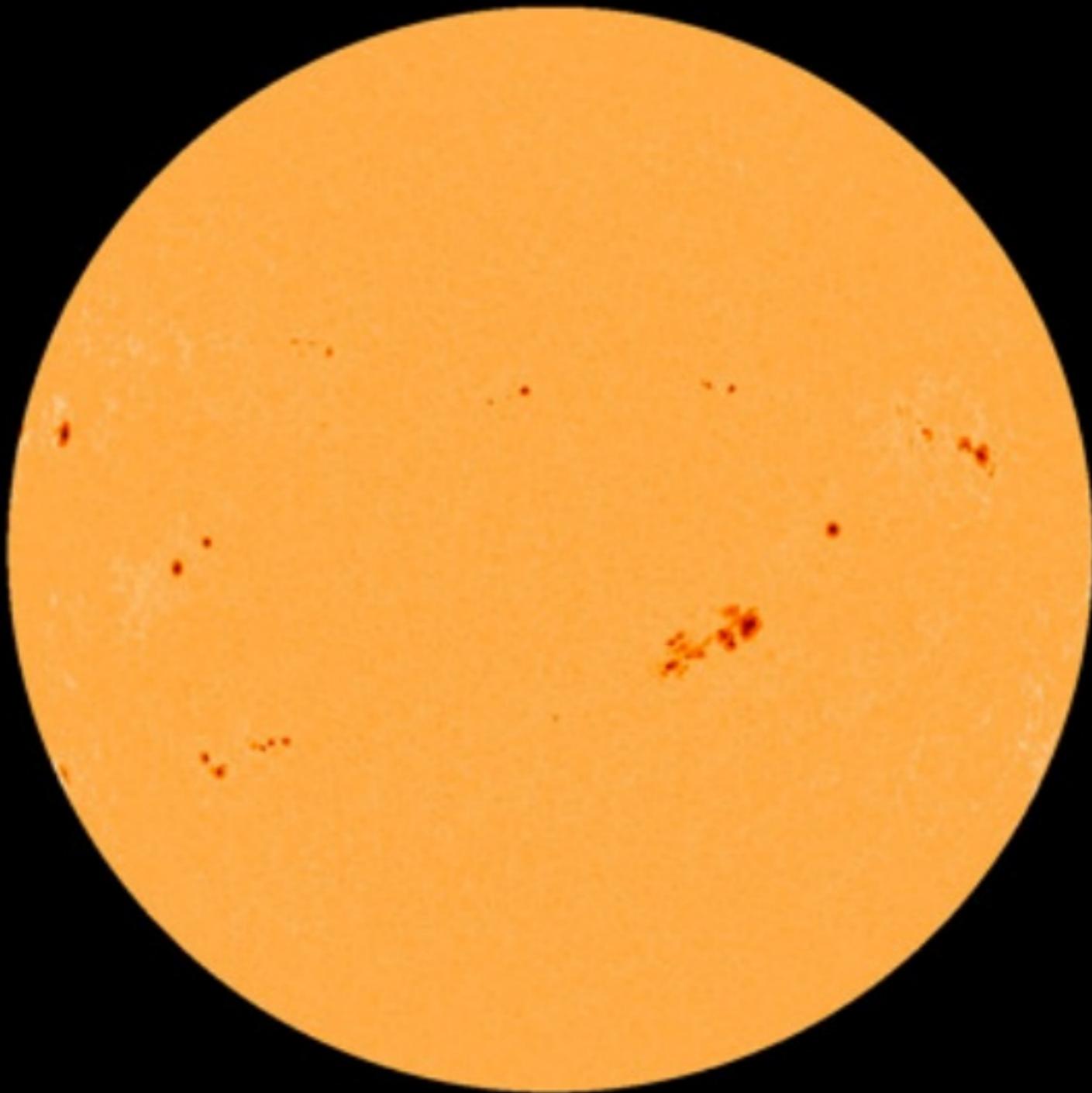


neurons

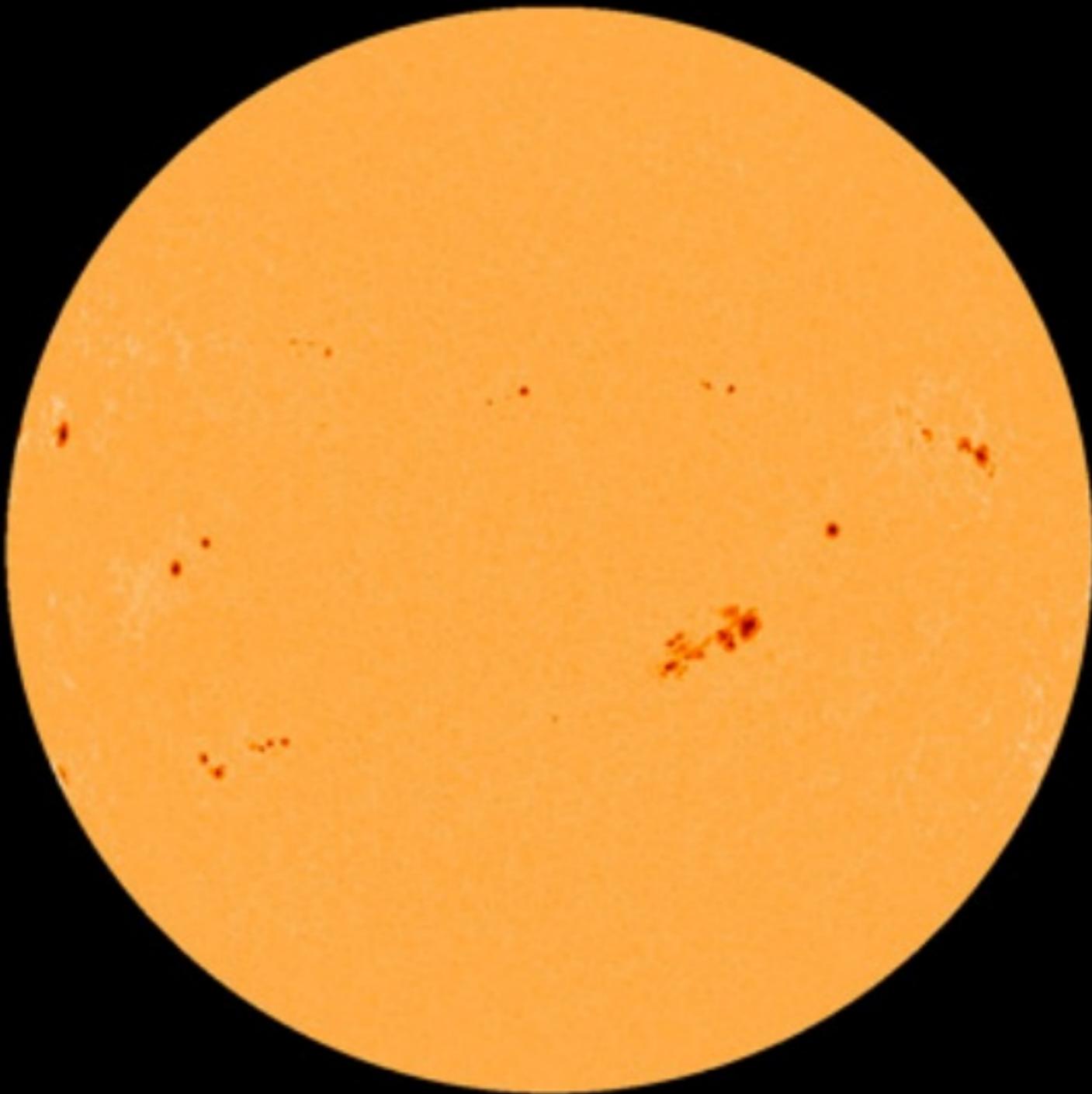


|           |          |             |
|-----------|----------|-------------|
| size      | $10^9$ m | $10^{-5}$ m |
| number    | 1        | $10^{11}$   |
| data size | 2 PB     | 12 TB/sec   |

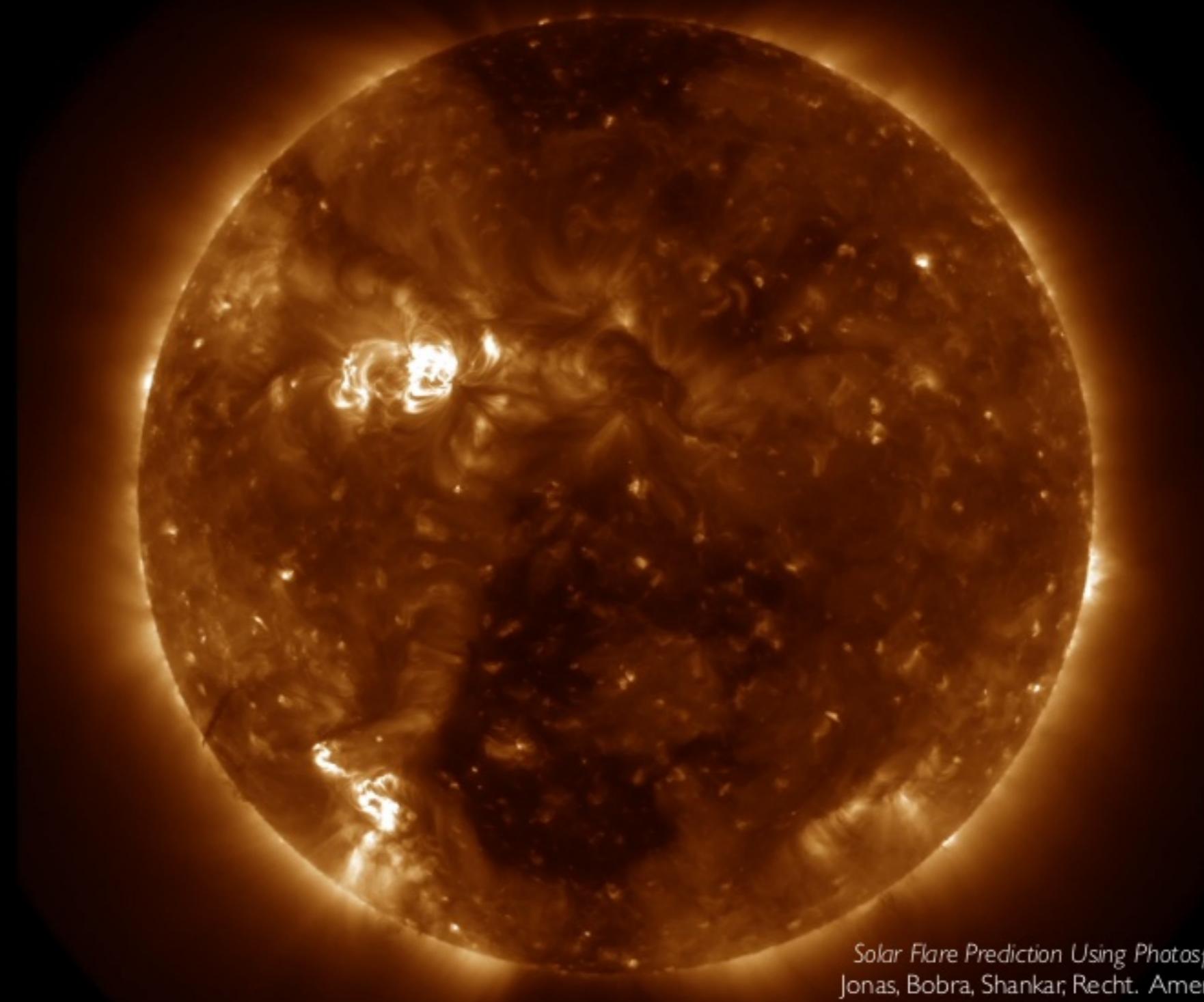
you are here .



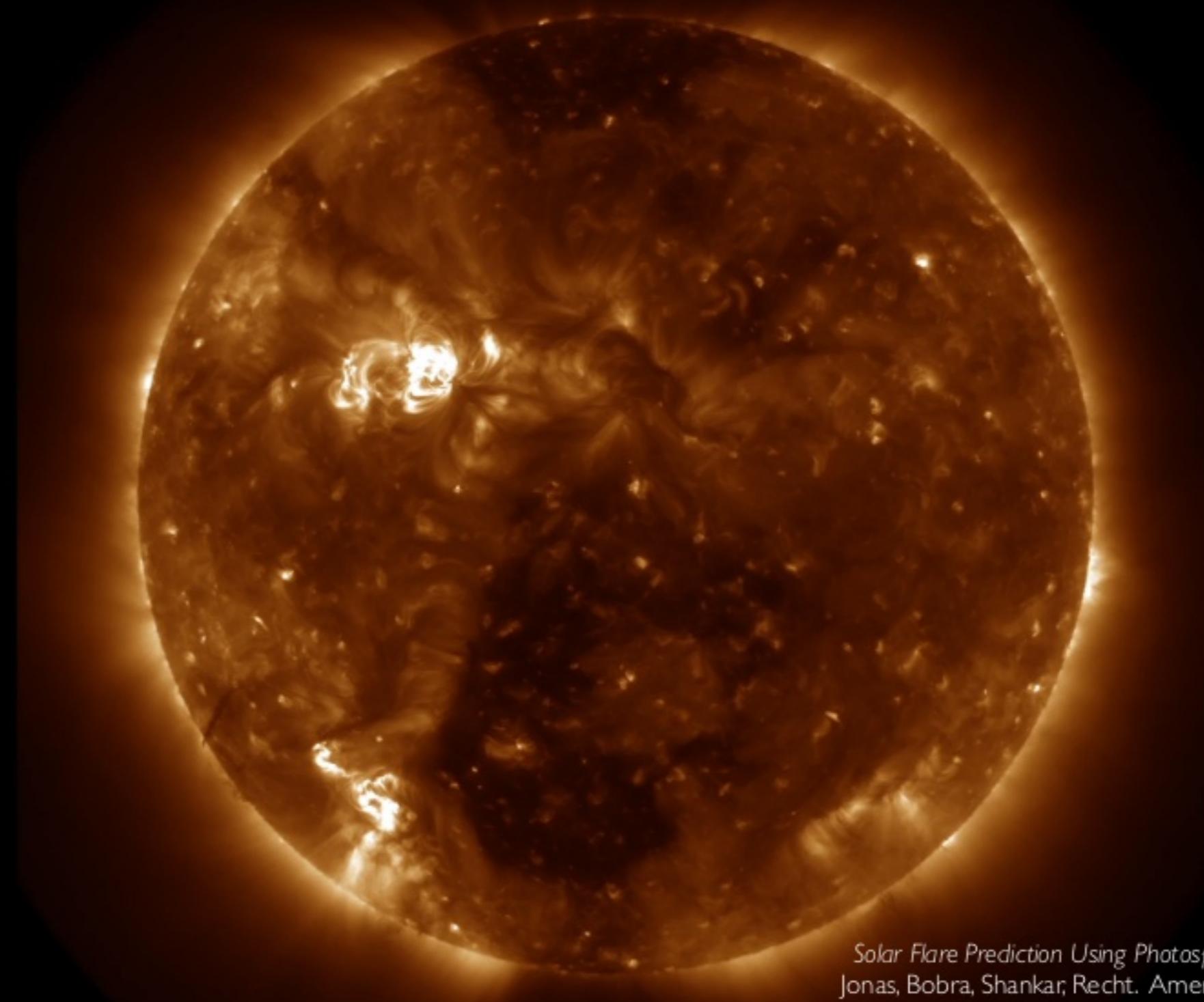
you are here .



Sun in UV (304 Å)



*Solar Flare Prediction Using Photospheric and Coronal Image Data.*  
Jonas, Bobra, Shankar; Recht. American Geophysical Union, 2016



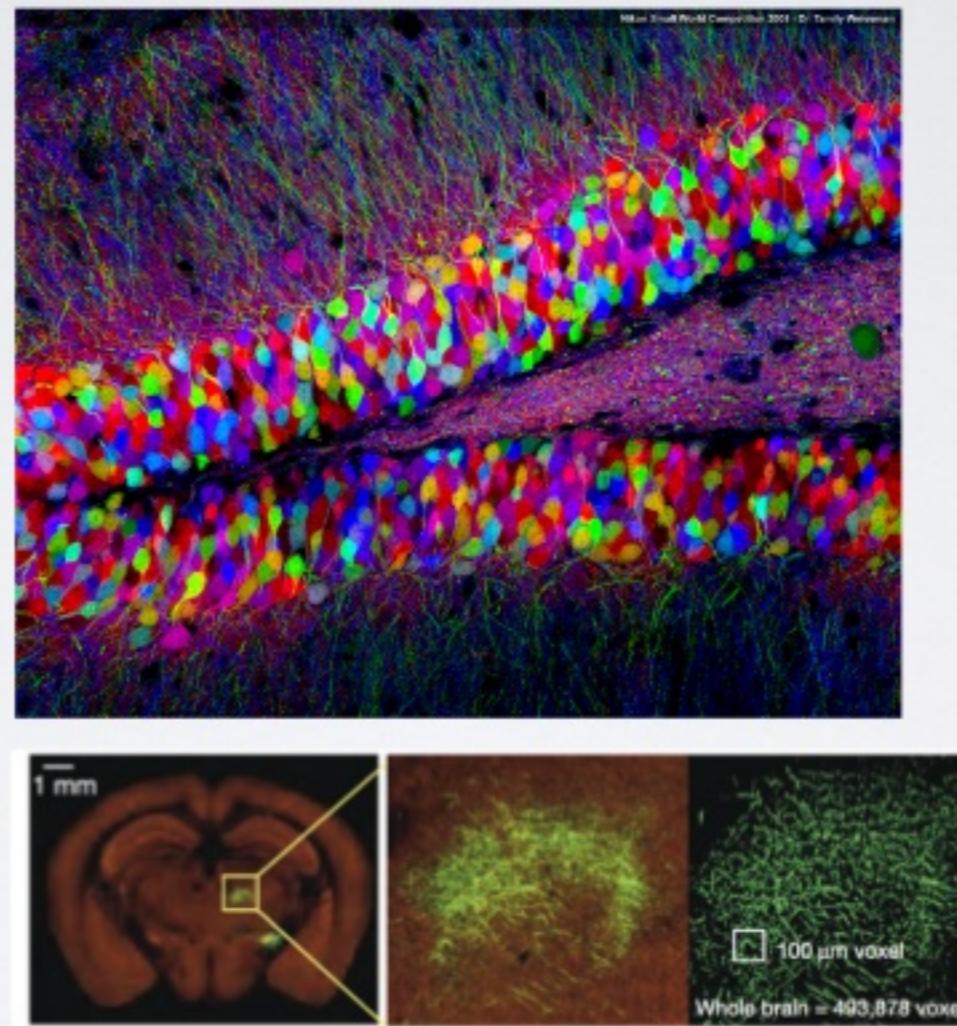
*Solar Flare Prediction Using Photospheric and Coronal Image Data.*  
Jonas, Bobra, Shankar; Recht. American Geophysical Union, 2016

# NEUROSCIENCE AT ALL SCALES

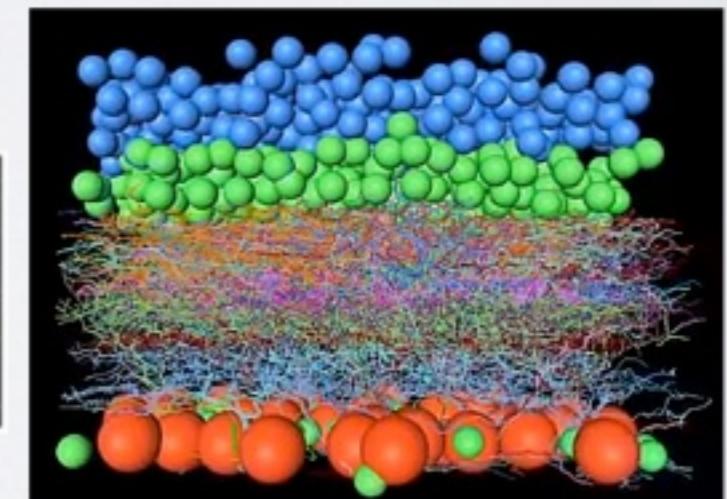
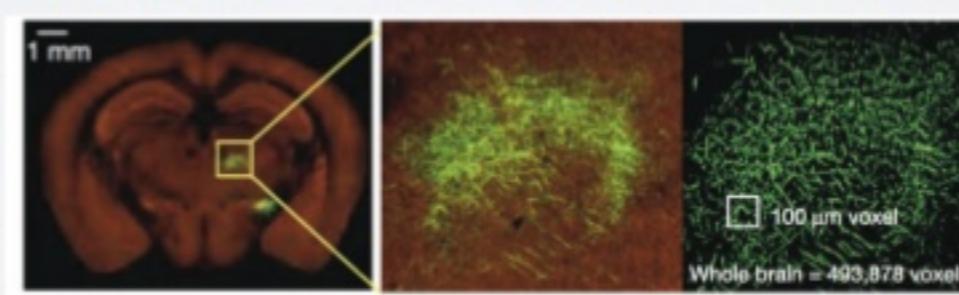
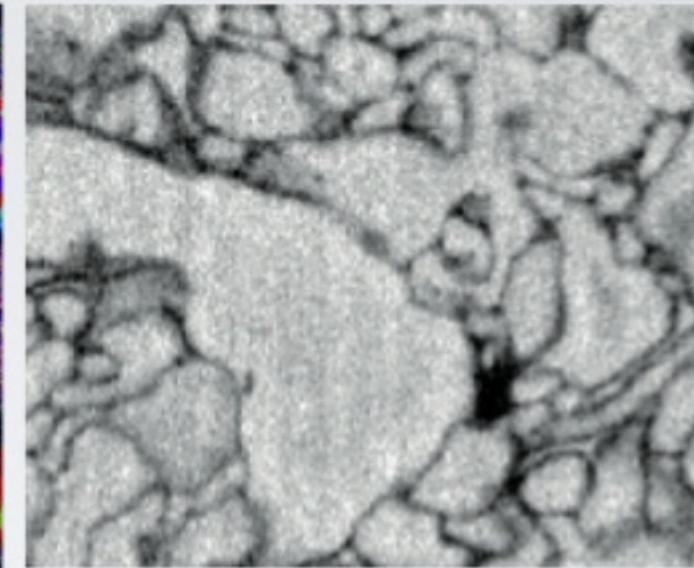
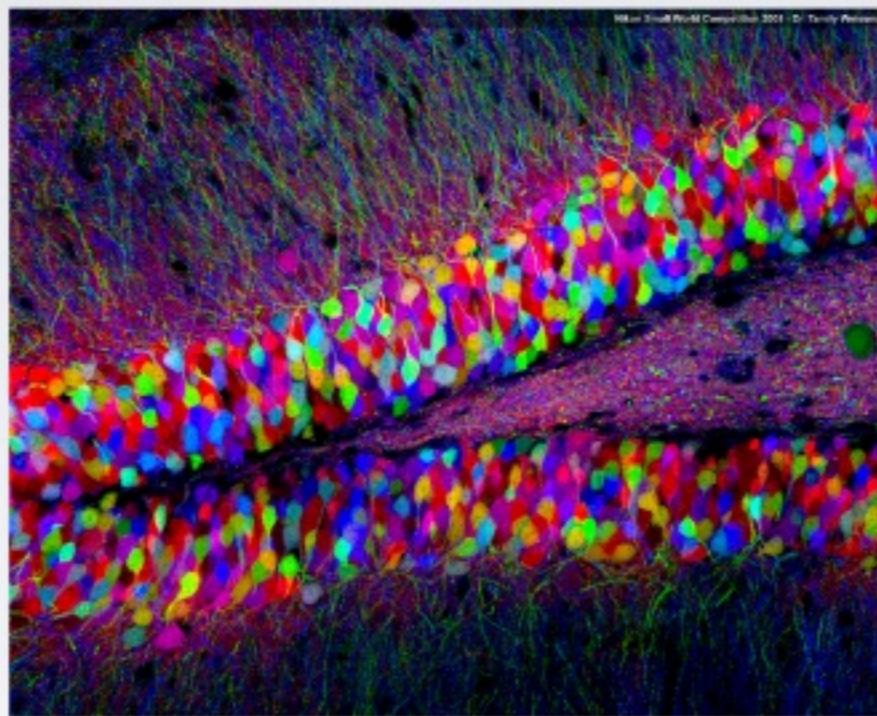
# NEUROSCIENCE AT ALL SCALES



# NEUROSCIENCE AT ALL SCALES



# NEUROSCIENCE AT ALL SCALES

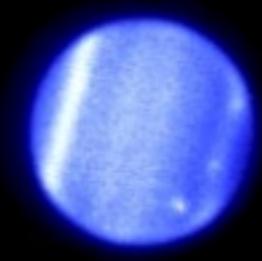




AO correction OFF



AO correction ON

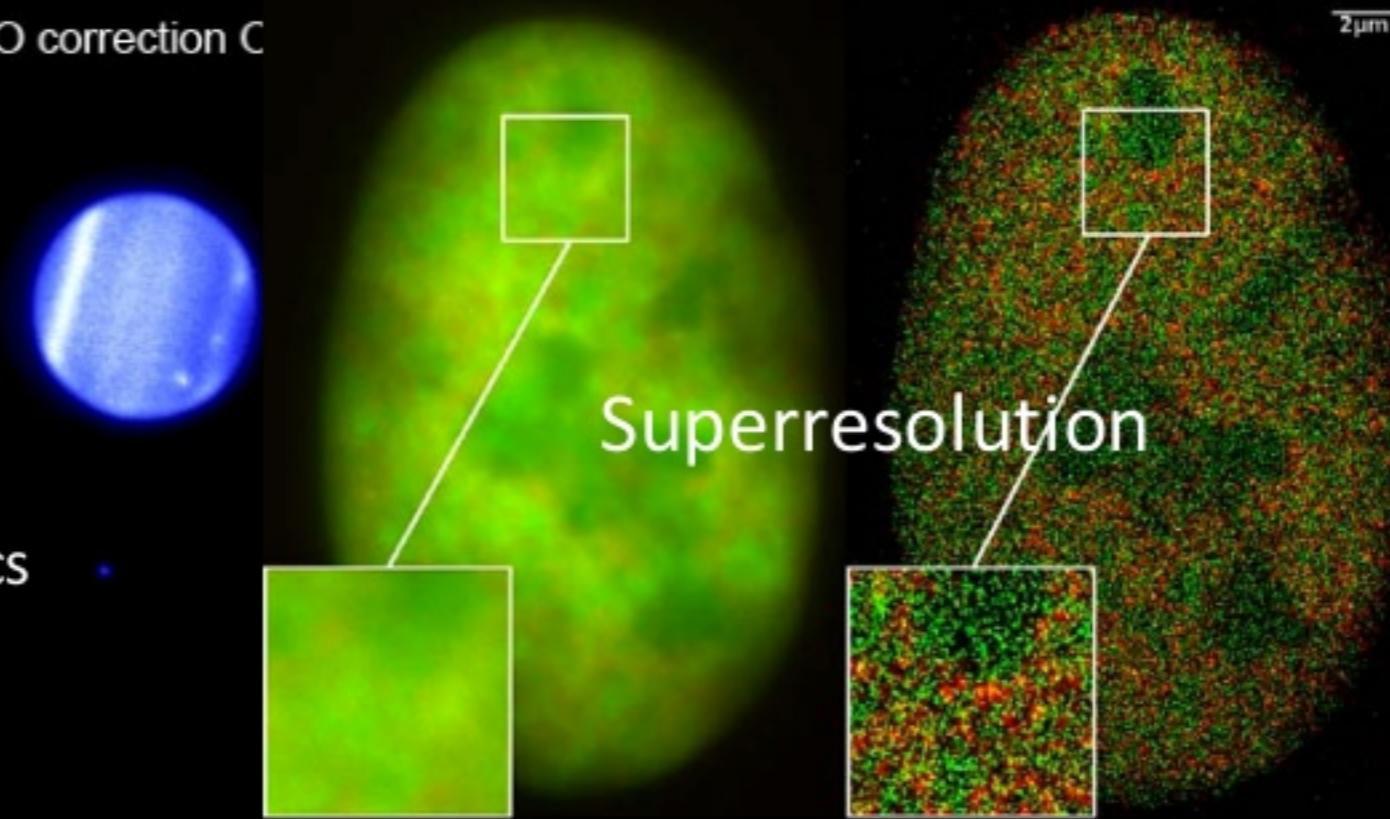


Adaptive Optics

AO correction OFF



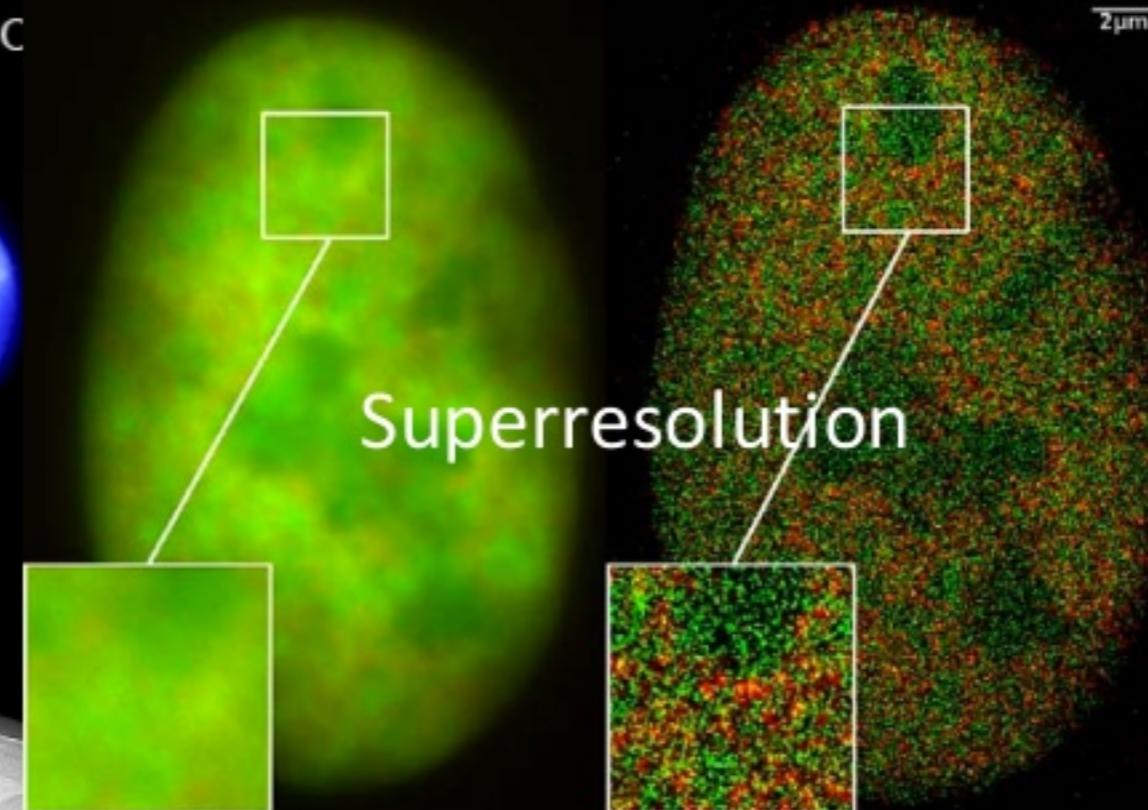
AO correction C



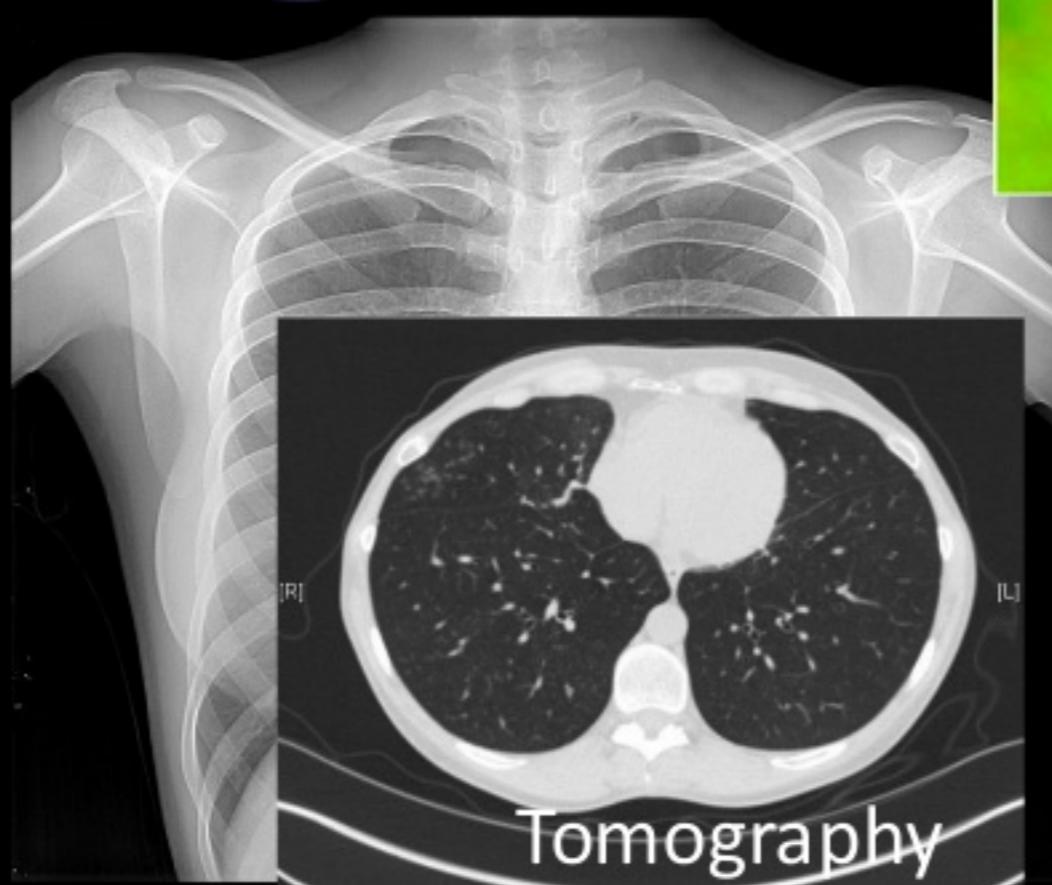
AO correction OFF



AO correction C



Adaptive Optics

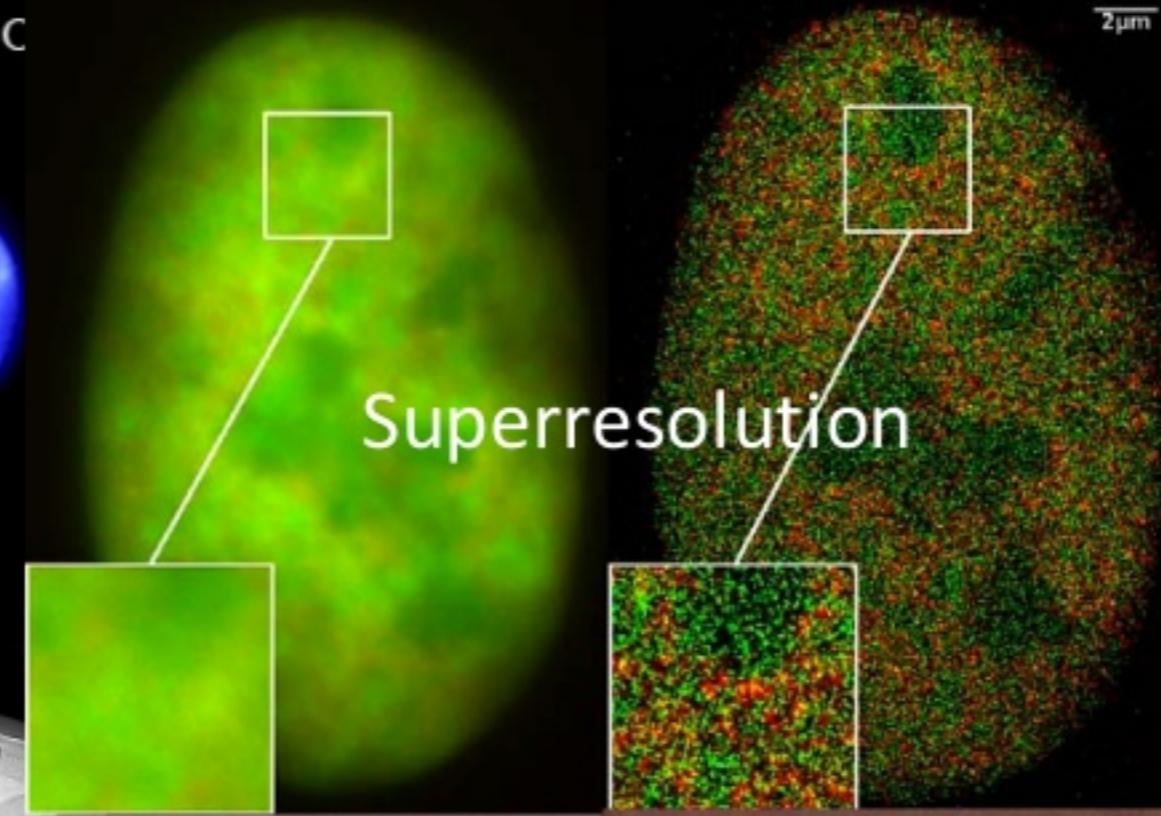


Tomography

AO correction OFF

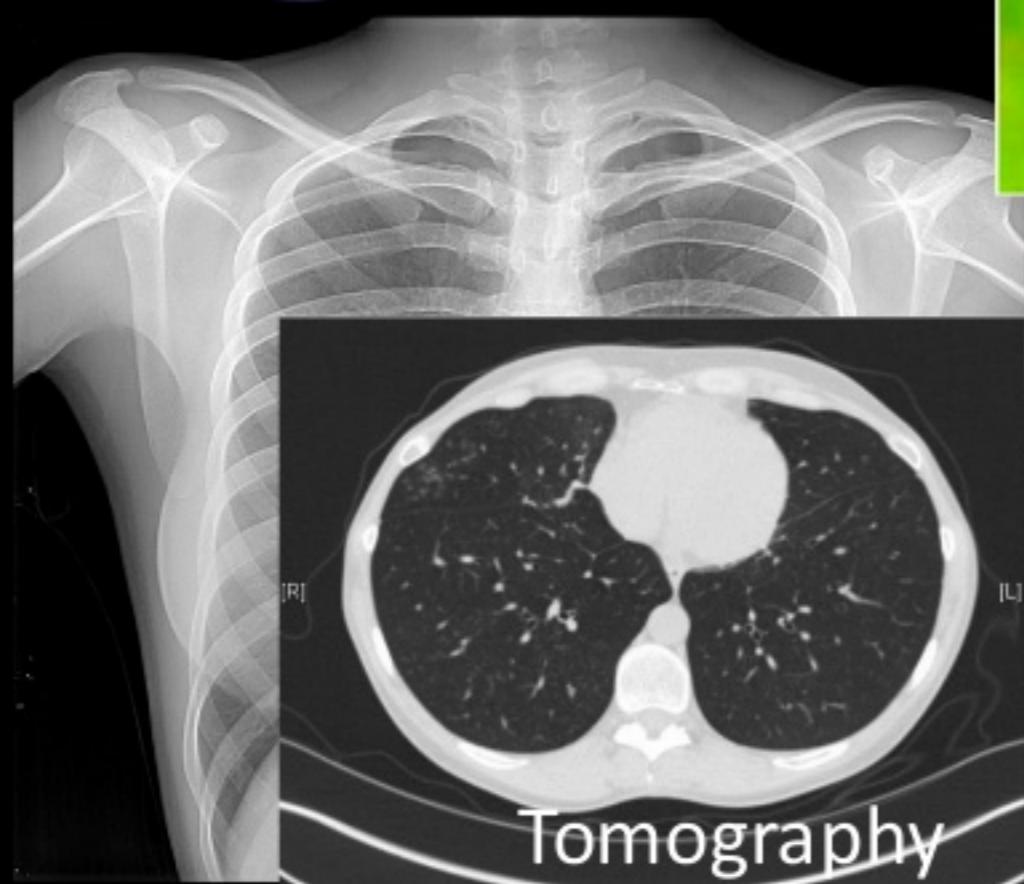


AO correction ON

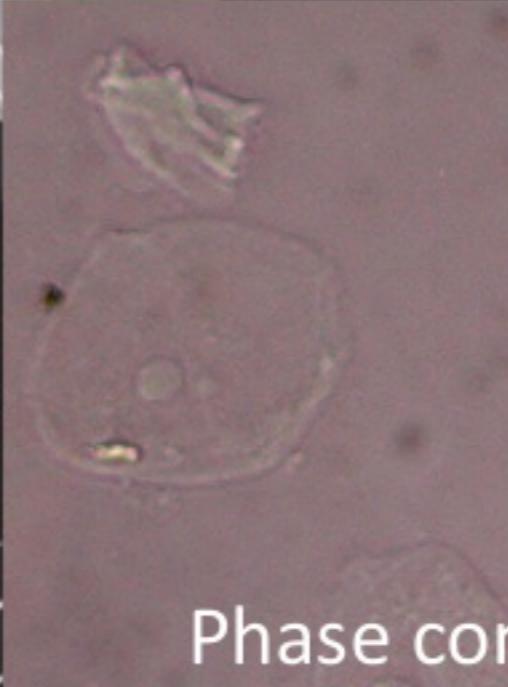


2 $\mu$ m

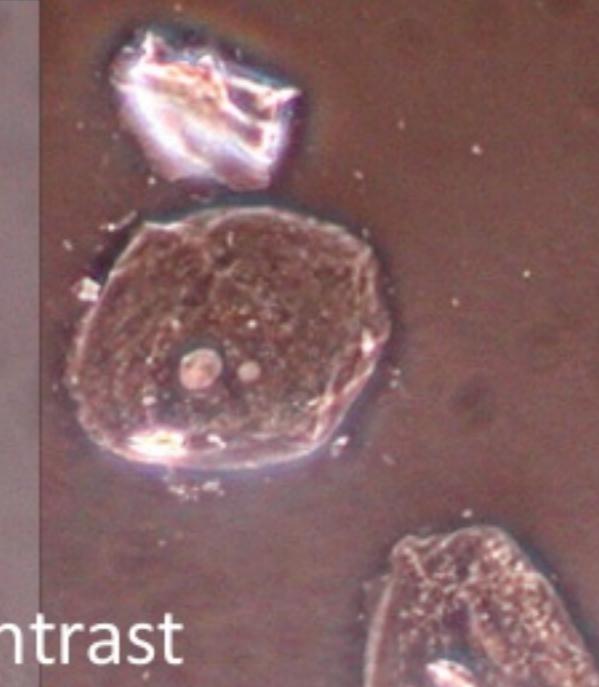
Adaptive Optics



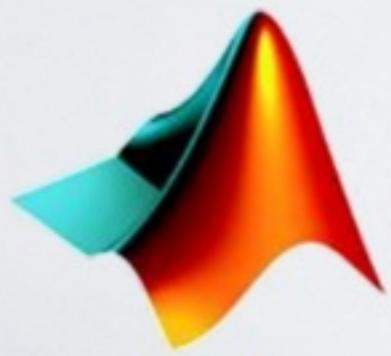
Tomography



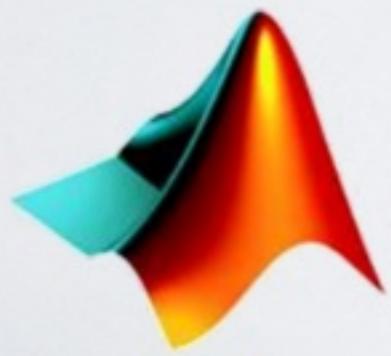
Phase contrast







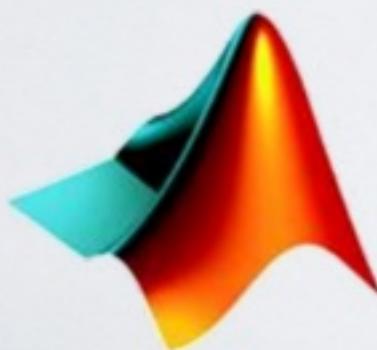
MATLAB®



MATLAB®



How do you get  
busy physicists and  
electrical engineers  
to give up Matlab?



MATLAB®

How do we get  
busy astronomers  
to give up IDL?

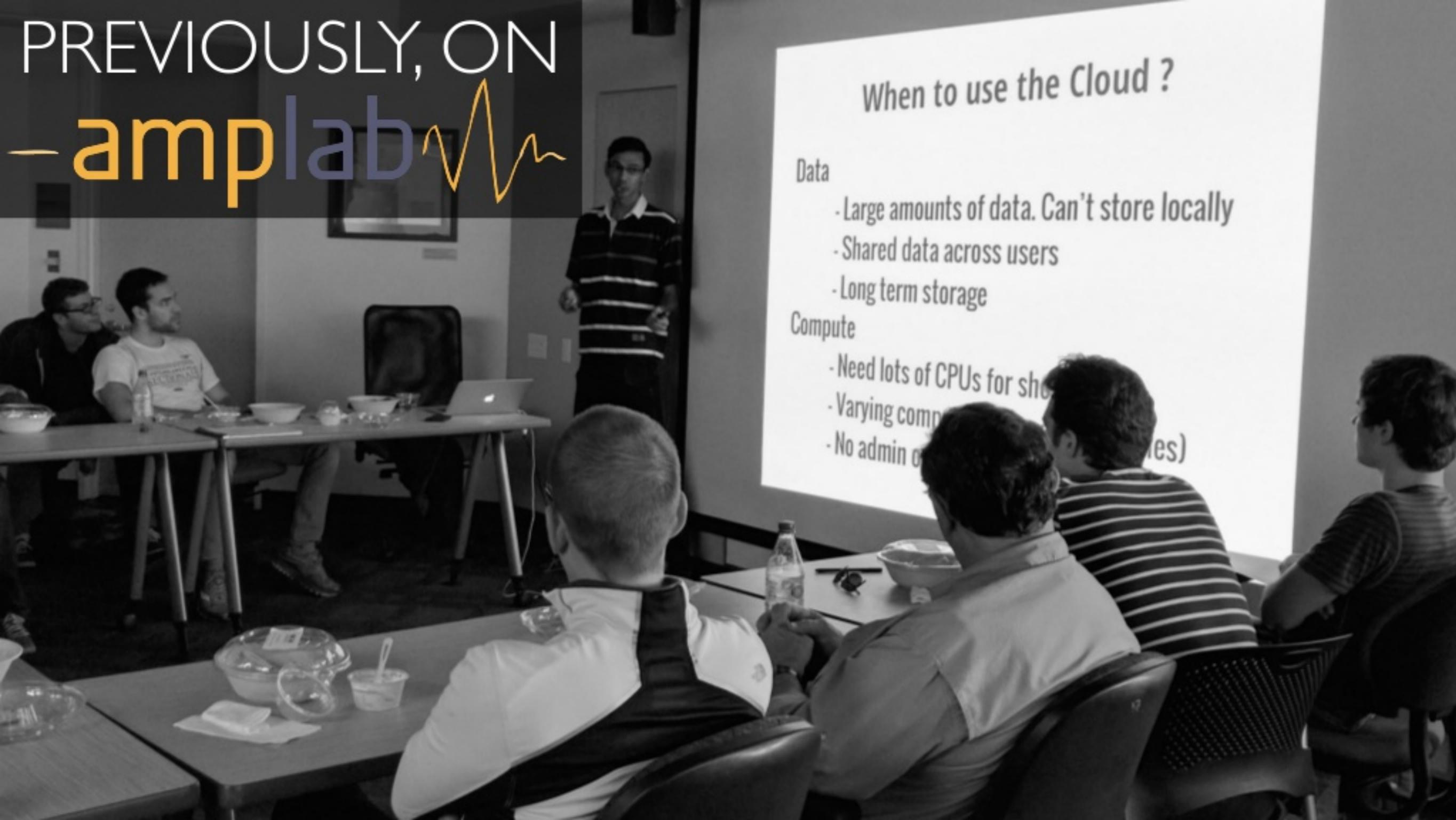


IDL®

PREVIOUSLY, ON  
- amplab



# PREVIOUSLY, ON - amplab



# PREVIOUSLY, ON - amplab



When to use the Cloud ?

Data

- Large amounts of data. Can't store locally
- Shared data across users
- Long term storage

Compute

- Need lots of CPUs for short time periods
- Varying compute requirements
- No admin overhead (no VMs)

# #THECLOUDISTOO DAMNHARD

| Amazon EC2 Instance Comparison |                          |                       |                          |                      |                              |                |            |               |           |            |               |             |                |                |                |
|--------------------------------|--------------------------|-----------------------|--------------------------|----------------------|------------------------------|----------------|------------|---------------|-----------|------------|---------------|-------------|----------------|----------------|----------------|
| Region                         |                          | Compute Power         |                          | Memory               |                              | Compute Skills |            | Storage       |           | Networking |               | Performance |                | Cost           |                |
| Region                         |                          | Compute Power         |                          | Memory               |                              | Compute Skills |            | Storage       |           | Networking |               | Performance |                | Cost           |                |
| Name                           | AMI Name                 | Memory                | Compute Skills           | Storage              | Bandwidth                    | Bandwidth      | Networking | Storage       | Bandwidth | Networking | Storage       | Performance | Cost           | Cost           | Cost           |
| cc2.8xlarge                    | cc2.8xlarge              | 38.4 GB               | 36 units                 | 16 vCPUs             | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.16xlarge                   | cc2.16xlarge             | 76.8 GB               | 72 units                 | 32 vCPUs             | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.32xlarge                   | cc2.32xlarge             | 153.6 GB              | 144 units                | 64 vCPUs             | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.64xlarge                   | cc2.64xlarge             | 307.2 GB              | 288 units                | 128 vCPUs            | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.128xlarge                  | cc2.128xlarge            | 614.4 GB              | 576 units                | 256 vCPUs            | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.256xlarge                  | cc2.256xlarge            | 122.88 GB             | 1152 units               | 512 vCPUs            | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.512xlarge                  | cc2.512xlarge            | 245.76 GB             | 2304 units               | 1024 vCPUs           | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.1024xlarge                 | cc2.1024xlarge           | 491.52 GB             | 4608 units               | 2048 vCPUs           | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.2048xlarge                 | cc2.2048xlarge           | 983.04 GB             | 9216 units               | 4096 vCPUs           | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.4096xlarge                 | cc2.4096xlarge           | 1966.08 GB            | 18432 units              | 8192 vCPUs           | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.8192xlarge                 | cc2.8192xlarge           | 3932.16 GB            | 36864 units              | 16384 vCPUs          | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.16384xlarge                | cc2.16384xlarge          | 7864.32 GB            | 73728 units              | 32768 vCPUs          | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.32768xlarge                | cc2.32768xlarge          | 15728.64 GB           | 147456 units             | 65536 vCPUs          | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.65536xlarge                | cc2.65536xlarge          | 31457.28 GB           | 294912 units             | 131072 vCPUs         | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.131072xlarge               | cc2.131072xlarge         | 62914.56 GB           | 589824 units             | 262144 vCPUs         | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.262144xlarge               | cc2.262144xlarge         | 125829.12 GB          | 1179648 units            | 524288 vCPUs         | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.524288xlarge               | cc2.524288xlarge         | 251658.24 GB          | 2359296 units            | 1048576 vCPUs        | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.1048576xlarge              | cc2.1048576xlarge        | 503316.48 GB          | 4718592 units            | 2097152 vCPUs        | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.2097152xlarge              | cc2.2097152xlarge        | 1006632.96 GB         | 9437184 units            | 4194304 vCPUs        | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.4194304xlarge              | cc2.4194304xlarge        | 2013265.92 GB         | 18874368 units           | 8388608 vCPUs        | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.8388608xlarge              | cc2.8388608xlarge        | 4026531.84 GB         | 37748736 units           | 16777216 vCPUs       | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.16777216xlarge             | cc2.16777216xlarge       | 8053063.68 GB         | 75497472 units           | 33554432 vCPUs       | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.33554432xlarge             | cc2.33554432xlarge       | 1610612.32 GB         | 150994944 units          | 67108864 vCPUs       | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.67108864xlarge             | cc2.67108864xlarge       | 3221224.64 GB         | 301989888 units          | 134217728 vCPUs      | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.134217728xlarge            | cc2.134217728xlarge      | 6442449.28 GB         | 603979536 units          | 268435456 vCPUs      | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.268435456xlarge            | cc2.268435456xlarge      | 12884898.56 GB        | 1207958880 units         | 536870912 vCPUs      | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.536870912xlarge            | cc2.536870912xlarge      | 25769797.12 GB        | 2415917600 units         | 1073741824 vCPUs     | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.1073741824xlarge           | cc2.1073741824xlarge     | 51539594.24 GB        | 4831835200 units         | 2147483648 vCPUs     | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.2147483648xlarge           | cc2.2147483648xlarge     | 103079188.48 GB       | 9663670400 units         | 4294967344 vCPUs     | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.4294967344xlarge           | cc2.4294967344xlarge     | 206158376.96 GB       | 19327340800 units        | 858993464 vCPUs      | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.858993464xlarge            | cc2.858993464xlarge      | 412316753.92 GB       | 38654681600 units        | 1717986928 vCPUs     | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.1717986928xlarge           | cc2.1717986928xlarge     | 824633507.84 GB       | 77309363200 units        | 3435973856 vCPUs     | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.3435973856xlarge           | cc2.3435973856xlarge     | 1649267015.68 GB      | 154618726400 units       | 6871947712 vCPUs     | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.6871947712xlarge           | cc2.6871947712xlarge     | 3298534031.36 GB      | 309237452800 units       | 1374389544 vCPUs     | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.1374389544xlarge           | cc2.1374389544xlarge     | 6597068062.72 GB      | 618474905600 units       | 274877888 vCPUs      | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.274877888xlarge            | cc2.274877888xlarge      | 13194136125.44 GB     | 1236949811200 units      | 549755776 vCPUs      | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.549755776xlarge            | cc2.549755776xlarge      | 26388272250.88 GB     | 2473899622400 units      | 1099511552 vCPUs     | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.1099511552xlarge           | cc2.1099511552xlarge     | 52776544501.76 GB     | 4947799244800 units      | 2199023040 vCPUs     | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.2199023040xlarge           | cc2.2199023040xlarge     | 10555308902.32 GB     | 9895598489600 units      | 4398046080 vCPUs     | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.4398046080xlarge           | cc2.4398046080xlarge     | 21110617804.64 GB     | 19791198979200 units     | 8796092160 vCPUs     | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.8796092160xlarge           | cc2.8796092160xlarge     | 42221235609.28 GB     | 39582397958400 units     | 17592184320 vCPUs    | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.17592184320xlarge          | cc2.17592184320xlarge    | 84442471218.56 GB     | 79164795916800 units     | 35184368640 vCPUs    | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.35184368640xlarge          | cc2.35184368640xlarge    | 168884942437.12 GB    | 158329591833600 units    | 70368737280 vCPUs    | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.70368737280xlarge          | cc2.70368737280xlarge    | 337769884874.24 GB    | 316659183667200 units    | 140735474560 vCPUs   | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.140735474560xlarge         | cc2.140735474560xlarge   | 675539769748.48 GB    | 633318367334400 units    | 281470949120 vCPUs   | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.281470949120xlarge         | cc2.281470949120xlarge   | 1351079539496.96 GB   | 1266636734668800 units   | 562941898240 vCPUs   | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.562941898240xlarge         | cc2.562941898240xlarge   | 2702159078993.92 GB   | 2533273469337600 units   | 1125883796480 vCPUs  | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.1125883796480xlarge        | cc2.1125883796480xlarge  | 5404318157987.84 GB   | 5066546938675200 units   | 2251767592960 vCPUs  | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.2251767592960xlarge        | cc2.2251767592960xlarge  | 10808636315975.68 GB  | 10133093877350400 units  | 4503535185920 vCPUs  | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.4503535185920xlarge        | cc2.4503535185920xlarge  | 21617272631951.36 GB  | 20266187754700800 units  | 9007070371840 vCPUs  | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.9007070371840xlarge        | cc2.9007070371840xlarge  | 43234545263852.72 GB  | 40532375509401600 units  | 18014140743680 vCPUs | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.18014140743680xlarge       | cc2.18014140743680xlarge | 86468690527705.44 GB  | 81064751018803200 units  | 36028281487360 vCPUs | 10000.0 MiB/s * 1000.0 MiB/s | 10 Gbit/s      | Low        | 10000.0 MiB/s | 10 Gbit/s | Low        | 10000.0 MiB/s | High        | \$1.000 hourly | \$1.000 hourly | \$1.000 hourly |
| cc2.36028281487360xlarge       | cc2.36028281487360xlarge | 172937381055410.88 GB | 162129502037606400 units | 72056562974720 vCPUs |                              |                |            |               |           |            |               |             |                |                |                |

# #THECLOUDISTOO DAMNHARD

- What type? what instance? What base image?

# #THECLOUDISTOO DAMNHARD

- What type? what instance? What base image?
- How many to spin up? What price? spot?

| EC2Instances.info Easy Amazon EC2 Instance Comparison   |              |                  |                      |                          |                              |                      |                     |                             |          |                      |                     |                        |                       |
|---|--------------|------------------|----------------------|--------------------------|------------------------------|----------------------|---------------------|-----------------------------|----------|----------------------|---------------------|------------------------|-----------------------|
| Region (US East vs. Region)   |              | Local Memory     |                      | Memory (1-16 GB options) |                              | Compute Units (vCPU) |                     | Bandwidth                   |          | Node                 |                     | Network Performance    |                       |
|   |              |                  |                      |                          |                              |                      |                     |                             |          |                      |                     |                        |                       |
| Name  | AMI Name     | Memory           | Compute Units (vCPU) | vCPUs                    | Bandwidth                    | Node                 | Network Performance | EBD Optimized Max Bandwidth | IFC Only | Linux On-Demand cost | Linux Reserved cost | Windows On-Demand cost | Windows Reserved cost |
| Cloud Compute Eight Extra Large   | ami-2f9e4e0c | 96.0 GB          | 16 units             | 16 vCPUs                 | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | 10 Gigabit          | 64 Gbps                     | No       | \$2.00/hourly        | \$1.00/hourly       | \$1.00/hourly          | \$1.00/hourly         |
| Cloud Compute Extra Large   | ami-2f9e4e0c | 60.0 GB          | 16 units             | 16 vCPUs                 | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | 10 Gigabit          | 64 Gbps                     | No       | \$1.00/hourly        | unreserved          | \$0.50/hourly          | unreserved            |
| t2.micro  | ami-2f9e4e0c | 1.0 GB           | 1 units              | 1 vCPU                   | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Low                 | 64 Gbps                     | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| t2.small  | ami-2f9e4e0c | 2.0 GB           | 2 units              | 2 vCPUs                  | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Low                 | 128 Gbps                    | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| t2.medium   | ami-2f9e4e0c | 4.0 GB           | 4 units              | 4 vCPUs                  | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Low                 | 256 Gbps                    | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| t2.large  | ami-2f9e4e0c | 8.0 GB           | 8 units              | 8 vCPUs                  | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Low                 | 512 Gbps                    | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| t2.xlarge   | ami-2f9e4e0c | 16.0 GB          | 16 units             | 16 vCPUs                 | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 1024 Gbps                   | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| t2.2xlarge  | ami-2f9e4e0c | 32.0 GB          | 32 units             | 32 vCPUs                 | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | High                | 2048 Gbps                   | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| t2.3xlarge  | ami-2f9e4e0c | 48.0 GB          | 48 units             | 48 vCPUs                 | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | High                | 3072 Gbps                   | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| t2.4xlarge  | ami-2f9e4e0c | 64.0 GB          | 64 units             | 64 vCPUs                 | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | High                | 4096 Gbps                   | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| c4 High CPU Large   | ami-2f9e4e0c | 1.0 GB           | 1 units              | 1 vCPU                   | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 64 Gbps                     | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| c4 High CPU Extra Large   | ami-2f9e4e0c | 16.0 GB          | 16 units             | 16 vCPUs                 | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | High                | 768 Gbps                    | Yes      | \$0.00/hourly        | \$0.10/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| c4t Extra Large   | ami-2f9e4e0c | 16.0 GB          | 16 units             | 16 vCPUs                 | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | High                | 1024 Gbps                   | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| c4t Quadruple Extra Large   | ami-2f9e4e0c | 32.0 GB          | 32 units             | 32 vCPUs                 | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | High                | 2048 Gbps                   | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| c4t Extra Extra Large   | ami-2f9e4e0c | 64.0 GB          | 64 units             | 64 vCPUs                 | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | High                | 4096 Gbps                   | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| c4 High CPU Large   | ami-2f9e4e0c | 1.0 GB           | 1 units              | 1 vCPU                   | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 64 Gbps                     | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| c4 High CPU Extra Large   | ami-2f9e4e0c | 16.0 GB          | 16 units             | 16 vCPUs                 | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | High                | 768 Gbps                    | Yes      | \$0.00/hourly        | \$0.10/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| c4t Extra Large   | ami-2f9e4e0c | 16.0 GB          | 16 units             | 16 vCPUs                 | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | High                | 1024 Gbps                   | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| c4t Quadruple Extra Large   | ami-2f9e4e0c | 32.0 GB          | 32 units             | 32 vCPUs                 | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | High                | 2048 Gbps                   | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| c4t Extra Extra Large   | ami-2f9e4e0c | 64.0 GB          | 64 units             | 64 vCPUs                 | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | High                | 4096 Gbps                   | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2 High Memory Large  | ami-2f9e4e0c | 32.0 GB          | 16 units             | 32 vCPUs                 | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 3072 Gbps                   | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2 High Memory Extra Large  | ami-2f9e4e0c | 64.0 GB          | 32 units             | 64 vCPUs                 | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 6144 Gbps                   | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Large   | ami-2f9e4e0c | 64.0 GB          | 32 units             | 64 vCPUs                 | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 6144 Gbps                   | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Extra Large   | ami-2f9e4e0c | 128.0 GB         | 64 units             | 128 vCPUs                | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 12288 Gbps                  | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Extra Extra Large   | ami-2f9e4e0c | 256.0 GB         | 128 units            | 256 vCPUs                | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 24576 Gbps                  | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Extra Extra Extra Large   | ami-2f9e4e0c | 512.0 GB         | 256 units            | 512 vCPUs                | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 49152 Gbps                  | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Extra Extra Extra Extra Large   | ami-2f9e4e0c | 1024.0 GB        | 512 units            | 1024 vCPUs               | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 98304 Gbps                  | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Extra Extra Extra Extra Extra Large   | ami-2f9e4e0c | 2048.0 GB        | 1024 units           | 2048 vCPUs               | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 196608 Gbps                 | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Extra Extra Extra Extra Extra Extra Large   | ami-2f9e4e0c | 4096.0 GB        | 2048 units           | 4096 vCPUs               | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 393216 Gbps                 | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Extra Extra Extra Extra Extra Extra Extra Large   | ami-2f9e4e0c | 8192.0 GB        | 4096 units           | 8192 vCPUs               | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 786432 Gbps                 | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Extra Extra Extra Extra Extra Extra Extra Extra Large   | ami-2f9e4e0c | 16384.0 GB       | 8192 units           | 16384 vCPUs              | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 1572864 Gbps                | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Large   | ami-2f9e4e0c | 32768.0 GB       | 16384 units          | 32768 vCPUs              | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 3145728 Gbps                | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Large   | ami-2f9e4e0c | 65536.0 GB       | 32768 units          | 65536 vCPUs              | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 6291456 Gbps                | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Large   | ami-2f9e4e0c | 131072.0 GB      | 65536 units          | 131072 vCPUs             | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 12582912 Gbps               | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Large   | ami-2f9e4e0c | 262144.0 GB      | 131072 units         | 262144 vCPUs             | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 25165824 Gbps               | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Large   | ami-2f9e4e0c | 524288.0 GB      | 262144 units         | 524288 vCPUs             | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 50331648 Gbps               | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Large   | ami-2f9e4e0c | 1048576.0 GB     | 524288 units         | 1048576 vCPUs            | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 100663296 Gbps              | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Large   | ami-2f9e4e0c | 2097152.0 GB     | 1048576 units        | 2097152 vCPUs            | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 201326592 Gbps              | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Large   | ami-2f9e4e0c | 4194304.0 GB     | 2097152 units        | 4194304 vCPUs            | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 402653184 Gbps              | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Large   | ami-2f9e4e0c | 8388608.0 GB     | 4194304 units        | 8388608 vCPUs            | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 805306368 Gbps              | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Large   | ami-2f9e4e0c | 16777216.0 GB    | 8388608 units        | 16777216 vCPUs           | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 1610612736 Gbps             | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Large   | ami-2f9e4e0c | 33554432.0 GB    | 16777216 units       | 33554432 vCPUs           | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 3221225472 Gbps             | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Large   | ami-2f9e4e0c | 67108864.0 GB    | 33554432 units       | 67108864 vCPUs           | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 6442450944 Gbps             | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Large   | ami-2f9e4e0c | 134217728.0 GB   | 67108864 units       | 134217728 vCPUs          | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 12884901888 Gbps            | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Large   | ami-2f9e4e0c | 268435456.0 GB   | 134217728 units      | 268435456 vCPUs          | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 25769803776 Gbps            | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Large   | ami-2f9e4e0c | 536870912.0 GB   | 268435456 units      | 536870912 vCPUs          | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 51539607552 Gbps            | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Large                                     | ami-2f9e4e0c | 1073741824.0 GB  | 536870912 units      | 1073741824 vCPUs         | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 103079215088 Gbps           | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Large                               | ami-2f9e4e0c | 2147483648.0 GB  | 1073741824 units     | 2147483648 vCPUs         | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 206158430176 Gbps           | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Large                         | ami-2f9e4e0c | 4294967296.0 GB  | 2147483648 units     | 4294967296 vCPUs         | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 412316860352 Gbps           | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Large                   | ami-2f9e4e0c | 8589934592.0 GB  | 4294967296 units     | 8589934592 vCPUs         | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 824633720704 Gbps           | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Large             | ami-2f9e4e0c | 17179869184.0 GB | 8589934592 units     | 17179869184 vCPUs        | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 1649267441408 Gbps          | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Large       | ami-2f9e4e0c | 34359738320.0 GB | 17179869184 units    | 34359738320 vCPUs        | 1000.0 Mbps (1 * 600.0 Mbps) | 64-bit               | Medium              | 3298534882816 Gbps          | Yes      | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          | \$0.00/hourly         |
| g2t Extra Large | ami-2f9e4e0c | 68719476640.0 GB | 34359738320 units    | 68719476640 vCPUs        | 100                          |                      |                     |                             |          |                      |                     |                        |                       |

# #THECLOUDISTOO DAMNHARD

- What type? what instance? What base image?
  - How many to spin up? What price? spot?
  - wait, Wait, WAIT oh god

# #THECLOUDISTOO DAMNHARD

- What type? what instance? What base image?
- How many to spin up? What price? spot?
- wait, Wait, WAIT oh god
- now what? DEVOPS

| EC2Instances.info Easy Amazon EC2 Instance Comparison |        |               |          |                              |           |                         |         |           |            |         |               |                     |               |                      |                     |                        |                       |
|---|--------|---------------|----------|------------------------------|-----------|-------------------------|---------|-----------|------------|---------|---------------|---------------------|---------------|----------------------|---------------------|------------------------|-----------------------|
| Region (US East vs. Region)                           |        | Local Memory  |          | Memory (1-16 GB) vs. No Swap |           | Compute Units (vCPU)    |         | Bandwidth |            | Node    |               | Network Performance |               |                      |                     |                        |                       |
| Region  | Region | Memory        | Swap     | Memory                       | Swap      | vCPUs                   | vCPUs   | Bandwidth | Nodes      | Network | Optimized     | Max Bandwidth       | IFC Only      | Linux On-Demand cost | Linux Reserved cost | Windows On-Demand cost | Windows Reserved cost |
| Compute Optimized Large                               |        | mt1.8xlarge   | 96.0 GB  | 16.0 units                   | 16 vCPUs  | 1000.0 (1.0 * 800.0 GB) | 80.0 GB | 10 Gbps   | 1x         | No      | \$2.00/hourly | \$1.00/hourly       | \$1.00/hourly | \$1.00/hourly        | \$1.00/hourly       | \$1.00/hourly          |                       |
| Compute Optimized Extra Large                         |        | mt2.8xlarge   | 96.0 GB  | 32.0 units                   | 16 vCPUs  | 1000.0 (1.0 * 800.0 GB) | 80.0 GB | 10 Gbps   | 1x         | No      | \$2.00/hourly | unavailable         | \$2.00/hourly | unavailable          | \$2.00/hourly       | unavailable            |                       |
| t2.micro  |        | t2.micro      | 1.0 GB   | 1.0 units                    | 1 vCPU    | 1.0 GB                  | 1.0 GB  | 1 Gbps    | 1x         | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| t2.small  |        | t2.small      | 2.0 GB   | 2.0 units                    | 2 vCPUs   | 2.0 GB                  | 2.0 GB  | 2 Gbps    | 1x         | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| t2.medium   |        | t2.medium     | 4.0 GB   | 4.0 units                    | 4 vCPUs   | 4.0 GB                  | 4.0 GB  | 4 Gbps    | 1x         | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| t2.large  |        | t2.large      | 8.0 GB   | 8.0 units                    | 8 vCPUs   | 8.0 GB                  | 8.0 GB  | 8 Gbps    | 1x         | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| t2.xlarge   |        | t2.xlarge     | 16.0 GB  | 16.0 units                   | 16 vCPUs  | 16.0 GB                 | 16.0 GB | 16 Gbps   | 1x         | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| m4.large  |        | m4.large      | 16.0 GB  | 16.0 units                   | 4 vCPUs   | 1.0 GB                  | 1.0 GB  | 1 Gbps    | 1x         | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| m4.xlarge   |        | m4.xlarge     | 32.0 GB  | 32.0 units                   | 8 vCPUs   | 2.0 GB                  | 2.0 GB  | 2 Gbps    | 1x         | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| m4.2xlarge  |        | m4.2xlarge    | 64.0 GB  | 64.0 units                   | 16 vCPUs  | 4.0 GB                  | 4.0 GB  | 4 Gbps    | 1x         | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| m4.4xlarge  |        | m4.4xlarge    | 128.0 GB | 128.0 units                  | 32 vCPUs  | 8.0 GB                  | 8.0 GB  | 8 Gbps    | 1x         | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| m4.8xlarge  |        | m4.8xlarge    | 256.0 GB | 256.0 units                  | 64 vCPUs  | 16.0 GB                 | 16.0 GB | 16 Gbps   | 1x         | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| c4.4xlarge  |        | c4.4xlarge    | 16.0 GB  | 16.0 units                   | 4 vCPUs   | 1.0 GB                  | 1.0 GB  | 1 Gbps    | 750.0 Mbps | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| c4.8xlarge  |        | c4.8xlarge    | 32.0 GB  | 32.0 units                   | 8 vCPUs   | 2.0 GB                  | 2.0 GB  | 2 Gbps    | 750.0 Mbps | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| c4.16xlarge   |        | c4.16xlarge   | 64.0 GB  | 64.0 units                   | 16 vCPUs  | 4.0 GB                  | 4.0 GB  | 4 Gbps    | 750.0 Mbps | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| c4.32xlarge   |        | c4.32xlarge   | 128.0 GB | 128.0 units                  | 32 vCPUs  | 8.0 GB                  | 8.0 GB  | 8 Gbps    | 750.0 Mbps | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| c4.64xlarge   |        | c4.64xlarge   | 256.0 GB | 256.0 units                  | 64 vCPUs  | 16.0 GB                 | 16.0 GB | 16 Gbps   | 750.0 Mbps | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| c5.4xlarge  |        | c5.4xlarge    | 16.0 GB  | 16.0 units                   | 4 vCPUs   | 1.0 GB                  | 1.0 GB  | 1 Gbps    | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| c5.8xlarge  |        | c5.8xlarge    | 32.0 GB  | 32.0 units                   | 8 vCPUs   | 2.0 GB                  | 2.0 GB  | 2 Gbps    | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| c5.16xlarge   |        | c5.16xlarge   | 64.0 GB  | 64.0 units                   | 16 vCPUs  | 4.0 GB                  | 4.0 GB  | 4 Gbps    | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| c5.32xlarge   |        | c5.32xlarge   | 128.0 GB | 128.0 units                  | 32 vCPUs  | 8.0 GB                  | 8.0 GB  | 8 Gbps    | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| c5.64xlarge   |        | c5.64xlarge   | 256.0 GB | 256.0 units                  | 64 vCPUs  | 16.0 GB                 | 16.0 GB | 16 Gbps   | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| c5.128xlarge  |        | c5.128xlarge  | 512.0 GB | 512.0 units                  | 128 vCPUs | 32.0 GB                 | 32.0 GB | 32 Gbps   | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| c5d.4xlarge   |        | c5d.4xlarge   | 16.0 GB  | 16.0 units                   | 4 vCPUs   | 1.0 GB                  | 1.0 GB  | 1 Gbps    | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| c5d.8xlarge   |        | c5d.8xlarge   | 32.0 GB  | 32.0 units                   | 8 vCPUs   | 2.0 GB                  | 2.0 GB  | 2 Gbps    | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| c5d.16xlarge  |        | c5d.16xlarge  | 64.0 GB  | 64.0 units                   | 16 vCPUs  | 4.0 GB                  | 4.0 GB  | 4 Gbps    | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| c5d.32xlarge  |        | c5d.32xlarge  | 128.0 GB | 128.0 units                  | 32 vCPUs  | 8.0 GB                  | 8.0 GB  | 8 Gbps    | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| c5d.64xlarge  |        | c5d.64xlarge  | 256.0 GB | 256.0 units                  | 64 vCPUs  | 16.0 GB                 | 16.0 GB | 16 Gbps   | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| c5d.128xlarge   |        | c5d.128xlarge | 512.0 GB | 512.0 units                  | 128 vCPUs | 32.0 GB                 | 32.0 GB | 32 Gbps   | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| g2.2xlarge  |        | g2.2xlarge    | 16.0 GB  | 16.0 units                   | 4 vCPUs   | 1.0 GB                  | 1.0 GB  | 1 Gbps    | 1 Gbps     | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| g2.8xlarge  |        | g2.8xlarge    | 32.0 GB  | 32.0 units                   | 8 vCPUs   | 2.0 GB                  | 2.0 GB  | 2 Gbps    | 1 Gbps     | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| g2.16xlarge   |        | g2.16xlarge   | 64.0 GB  | 64.0 units                   | 16 vCPUs  | 4.0 GB                  | 4.0 GB  | 4 Gbps    | 1 Gbps     | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| g2.32xlarge   |        | g2.32xlarge   | 128.0 GB | 128.0 units                  | 32 vCPUs  | 8.0 GB                  | 8.0 GB  | 8 Gbps    | 1 Gbps     | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| g3.4xlarge  |        | g3.4xlarge    | 16.0 GB  | 16.0 units                   | 4 vCPUs   | 1.0 GB                  | 1.0 GB  | 1 Gbps    | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| g3.8xlarge  |        | g3.8xlarge    | 32.0 GB  | 32.0 units                   | 8 vCPUs   | 2.0 GB                  | 2.0 GB  | 2 Gbps    | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| g3.16xlarge   |        | g3.16xlarge   | 64.0 GB  | 64.0 units                   | 16 vCPUs  | 4.0 GB                  | 4.0 GB  | 4 Gbps    | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| g3.32xlarge   |        | g3.32xlarge   | 128.0 GB | 128.0 units                  | 32 vCPUs  | 8.0 GB                  | 8.0 GB  | 8 Gbps    | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| g3.64xlarge   |        | g3.64xlarge   | 256.0 GB | 256.0 units                  | 64 vCPUs  | 16.0 GB                 | 16.0 GB | 16 Gbps   | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| g3.128xlarge  |        | g3.128xlarge  | 512.0 GB | 512.0 units                  | 128 vCPUs | 32.0 GB                 | 32.0 GB | 32 Gbps   | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| g4.4xlarge  |        | g4.4xlarge    | 16.0 GB  | 16.0 units                   | 4 vCPUs   | 1.0 GB                  | 1.0 GB  | 1 Gbps    | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| g4.8xlarge  |        | g4.8xlarge    | 32.0 GB  | 32.0 units                   | 8 vCPUs   | 2.0 GB                  | 2.0 GB  | 2 Gbps    | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| g4.16xlarge   |        | g4.16xlarge   | 64.0 GB  | 64.0 units                   | 16 vCPUs  | 4.0 GB                  | 4.0 GB  | 4 Gbps    | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| g4.32xlarge   |        | g4.32xlarge   | 128.0 GB | 128.0 units                  | 32 vCPUs  | 8.0 GB                  | 8.0 GB  | 8 Gbps    | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| g4.64xlarge   |        | g4.64xlarge   | 256.0 GB | 256.0 units                  | 64 vCPUs  | 16.0 GB                 | 16.0 GB | 16 Gbps   | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| g4.128xlarge  |        | g4.128xlarge  | 512.0 GB | 512.0 units                  | 128 vCPUs | 32.0 GB                 | 32.0 GB | 32 Gbps   | 10 Gbps    | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| i3.4xlarge  |        | i3.4xlarge    | 16.0 GB  | 16.0 units                   | 4 vCPUs   | 1.0 GB                  | 1.0 GB  | 1 Gbps    | 1 Gbps     | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| i3.8xlarge  |        | i3.8xlarge    | 32.0 GB  | 32.0 units                   | 8 vCPUs   | 2.0 GB                  | 2.0 GB  | 2 Gbps    | 1 Gbps     | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| i3.16xlarge   |        | i3.16xlarge   | 64.0 GB  | 64.0 units                   | 16 vCPUs  | 4.0 GB                  | 4.0 GB  | 4 Gbps    | 1 Gbps     | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| i3.32xlarge   |        | i3.32xlarge   | 128.0 GB | 128.0 units                  | 32 vCPUs  | 8.0 GB                  | 8.0 GB  | 8 Gbps    | 1 Gbps     | Yes     | \$0.00/hourly | \$0.00/hourly       | \$0.00/hourly | \$0.00/hourly        | \$0.00/hourly       | \$0.00/hourly          |                       |
| i3.64xlarge   |        | i3.64xlarge   | 256.0 GB | 256.0 units                  | 64 vCPUs  | 16.0 GB                 | 16.0 GB | 16        |            |         |               |                     |               |                      |                     |                        |                       |

# WHAT DO WE WANT?

## I. **Very little overhead for setup**

once someone has an AWS account. In particular,  
no persistent overhead -- you don't have to keep  
a large (expensive) cluster up and you don't have  
to wait 10+ min for a cluster to come up

# WHAT DO WE WANT?

**2.** As close to zero overhead for users as possible

In particular, **anyone who can write python** should be able to invoke it through a reasonable interface. It should support all legacy code

# WHAT DO WE WANT?

3. Target jobs that run in the **minutes-or-more regime**.



PyWren



[pywren.io](http://pywren.io)

“Most wrens are small and rather inconspicuous, except  
for their loud and often complex songs.”

pip install pywren

# PYWREN:THE API

```
import pywren
import numpy as np

def addone(x):
    return x + 1

wrenexec = pywren.default_executor()
xlist = np.arange(10)
futures = wrenexec.map(addone, xlist)

print [f.result() for f in futures]
```

The output is as expected:

```
[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
```

USING  
“SERVERLESS INFRASTRUCTURE”



**ANACONDA®**



Powered by Continuum Analytics



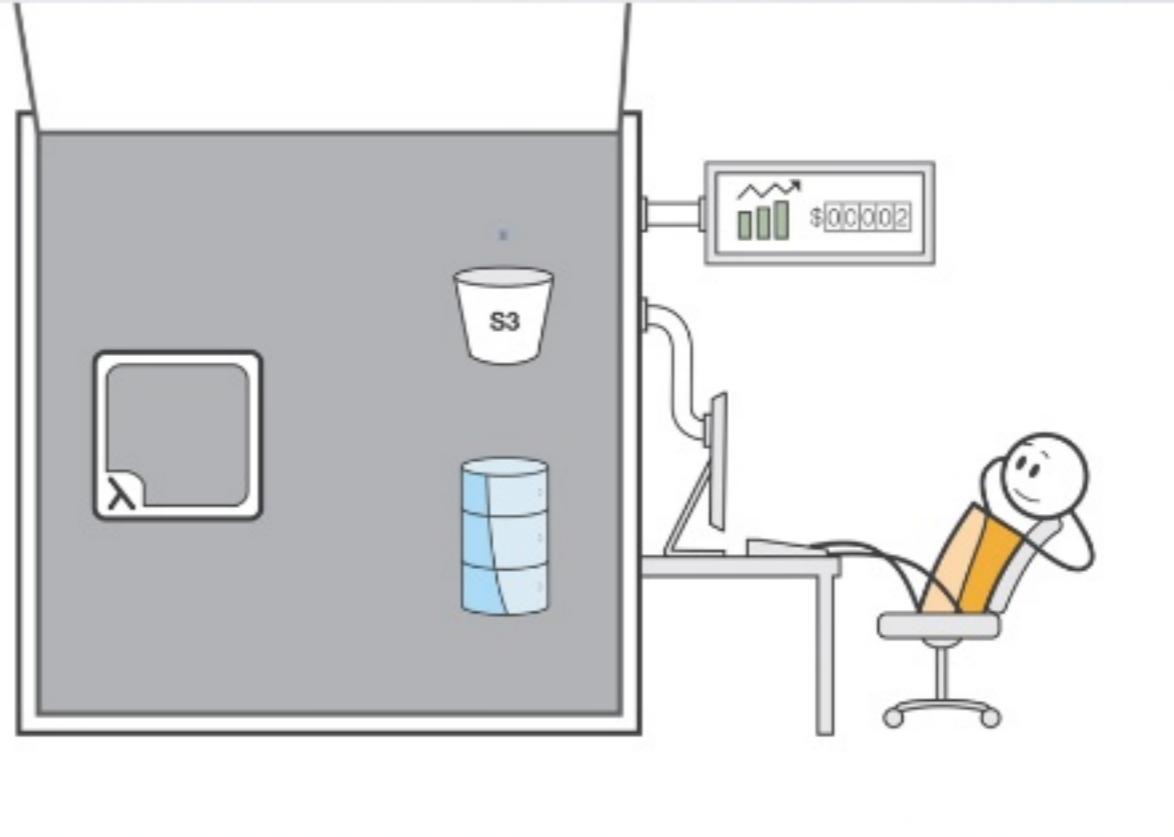
AWS Lambda

Run code without thinking about servers.  
Pay for only the compute time you consume.

[Get started with AWS Lambda](#)

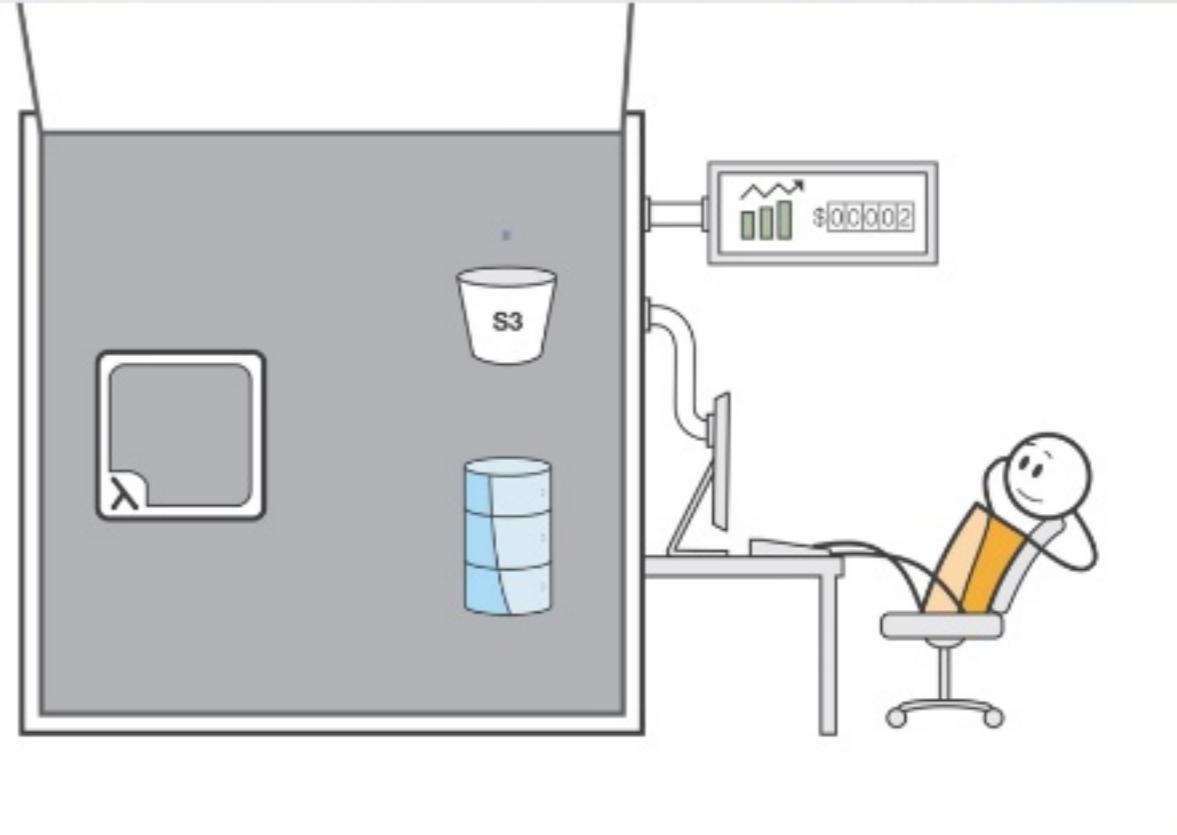
# AWS LAMBDA

- 300 seconds  
single-core (AVX2)
- 512 MB in /tmp
- 1.5GB RAM
- Python, Java, Node



# AWS LAMBDA

- 300 seconds  
single-core (AVX2)
- 512 MB in /tmp
- 1.5GB RAM
- Python, Java, Node



# AWS LAMBDA

- 300 seconds single-core (AVX2)
- 512 MB in /tmp
- 1.5GB RAM
- Python, Java, Node



Google Cloud Platform

CLOUD FUNCTIONS ALPHA

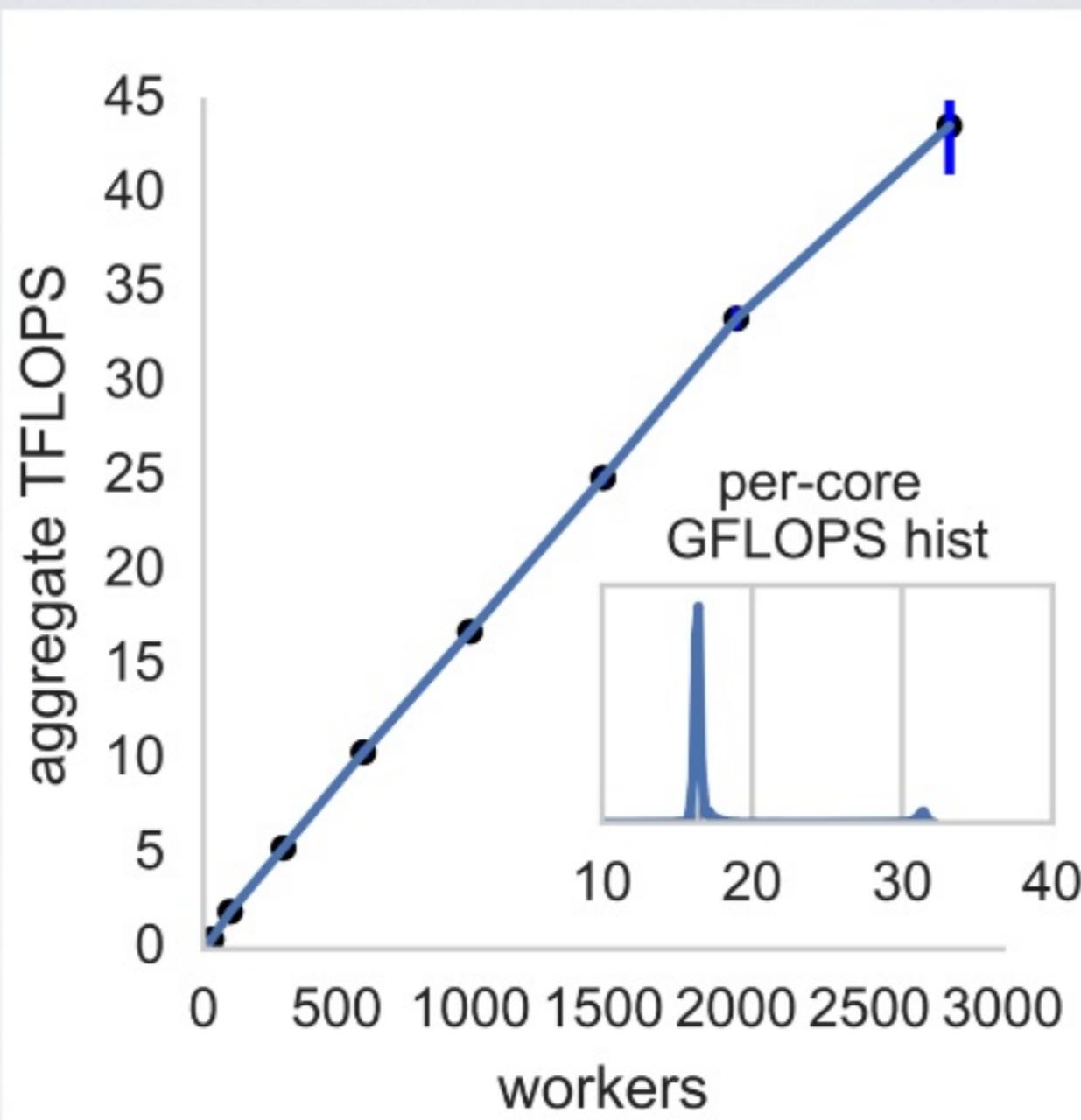
A serverless platform for building event-based microservices

Microsoft Azure

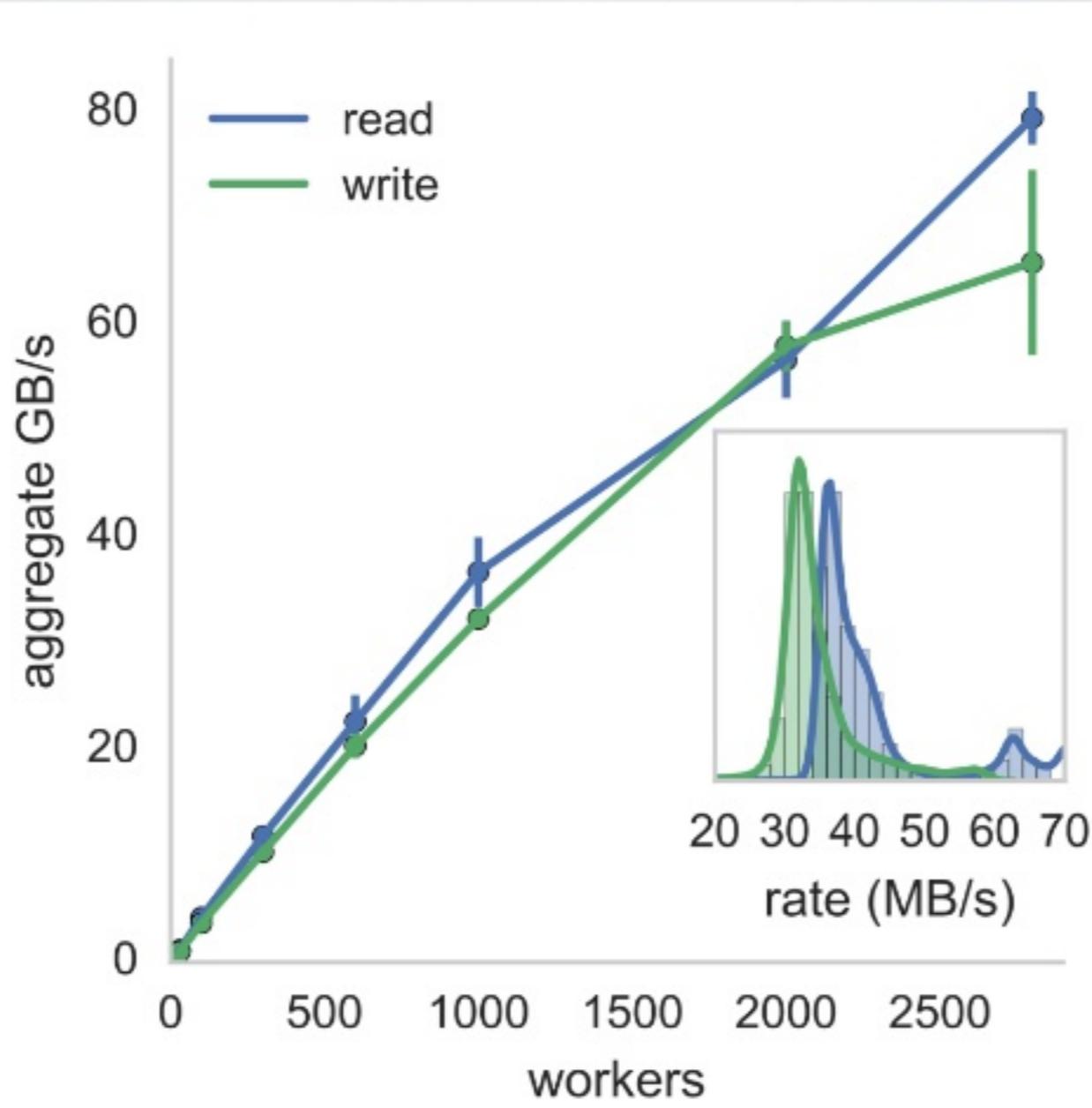
Azure Functions

Process events with a serverless code architecture

# LAMBDA SCALABILITY



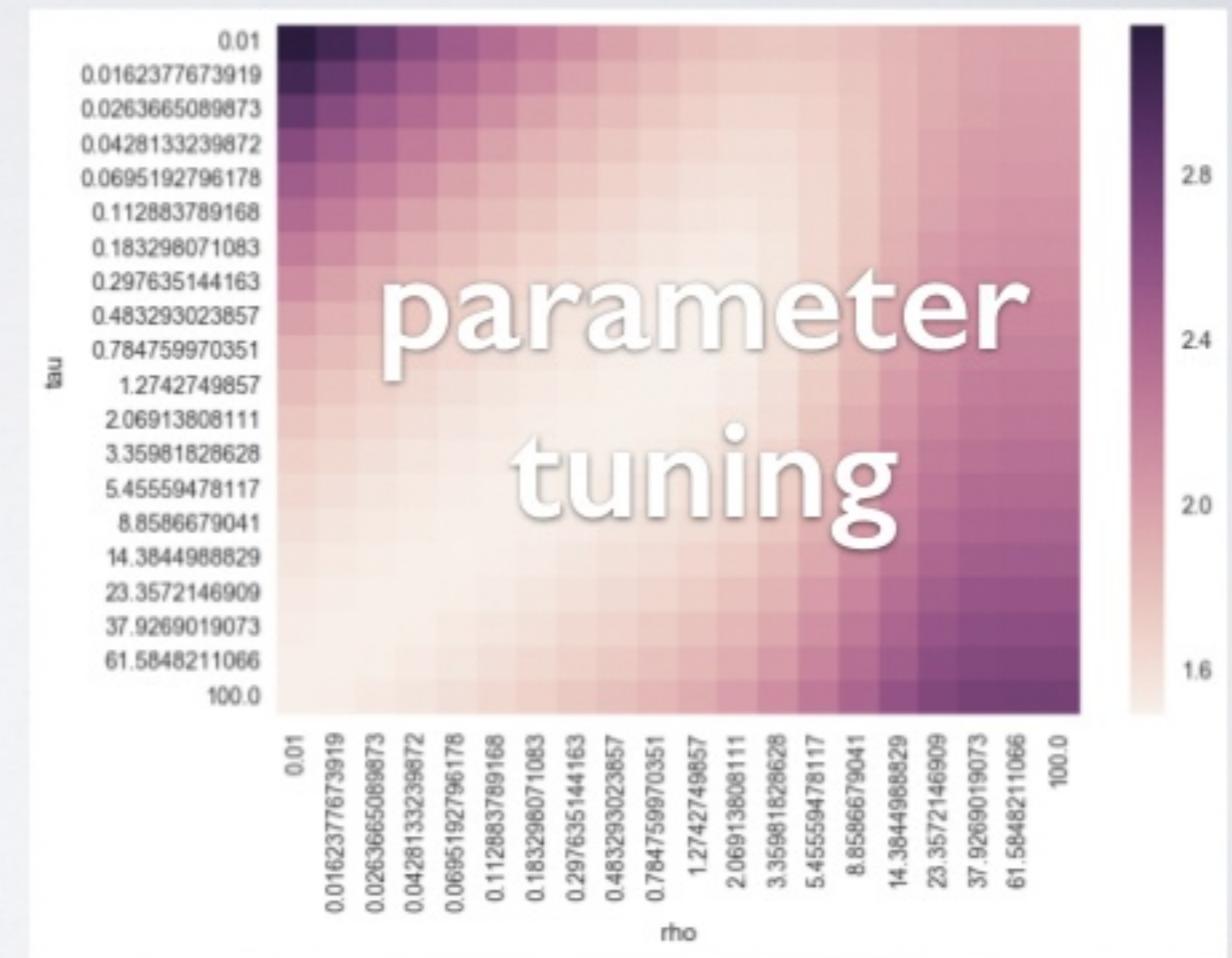
# LAMBDA SCALABILITY

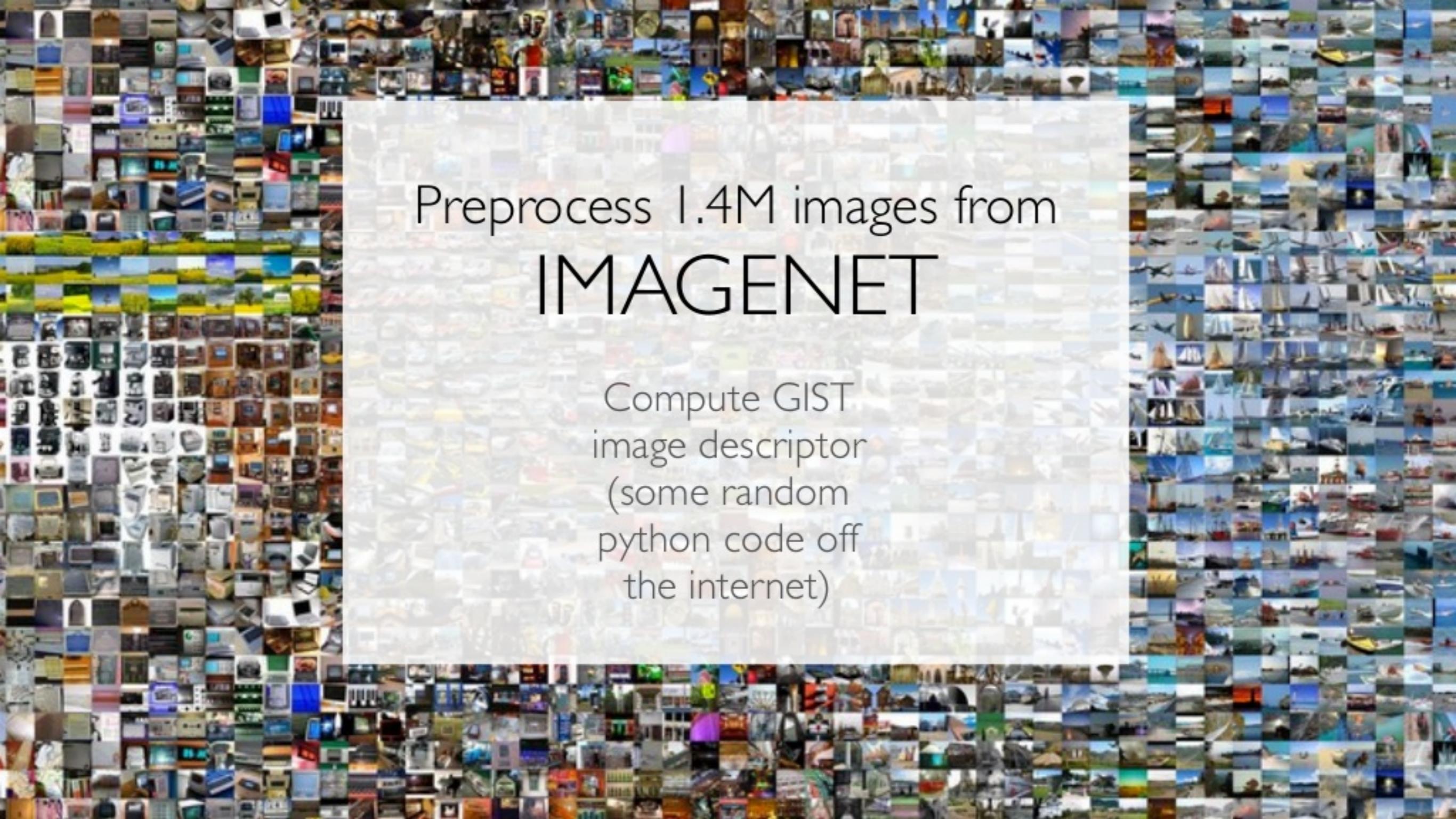


YOU CAN DO A LOT OF  
WORK WITH MAP!



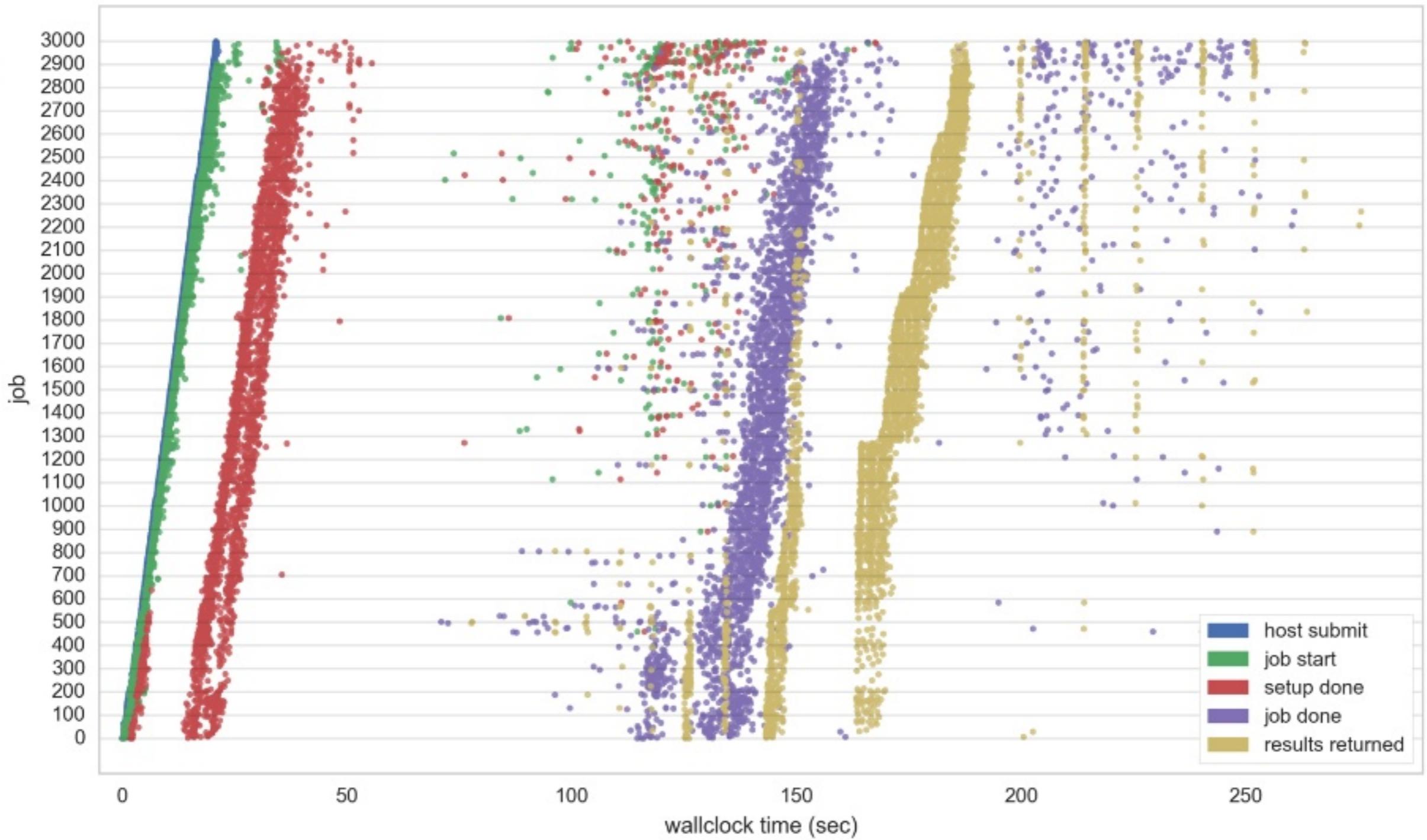
# YOU CAN DO A LOT OF WORK WITH MAP!

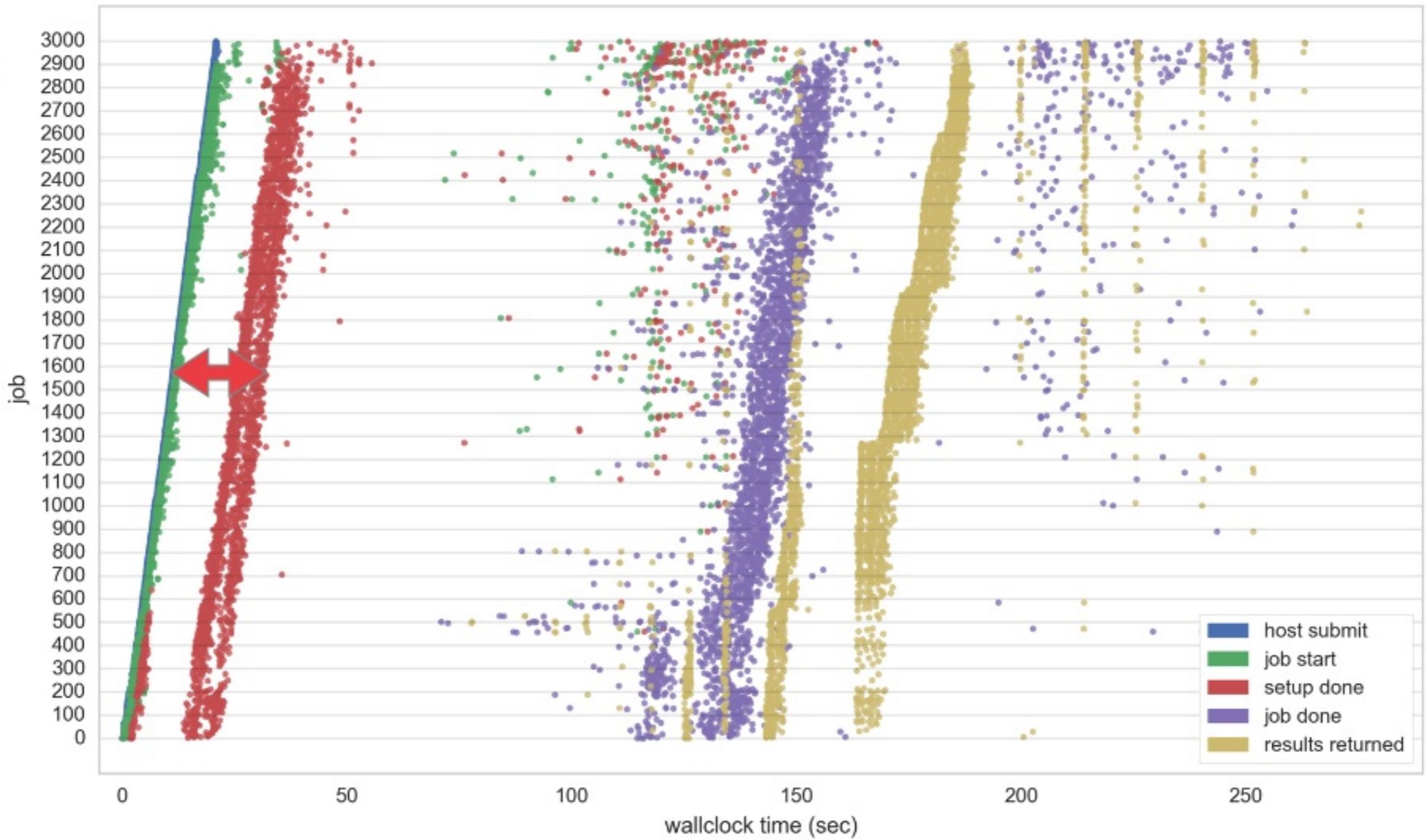


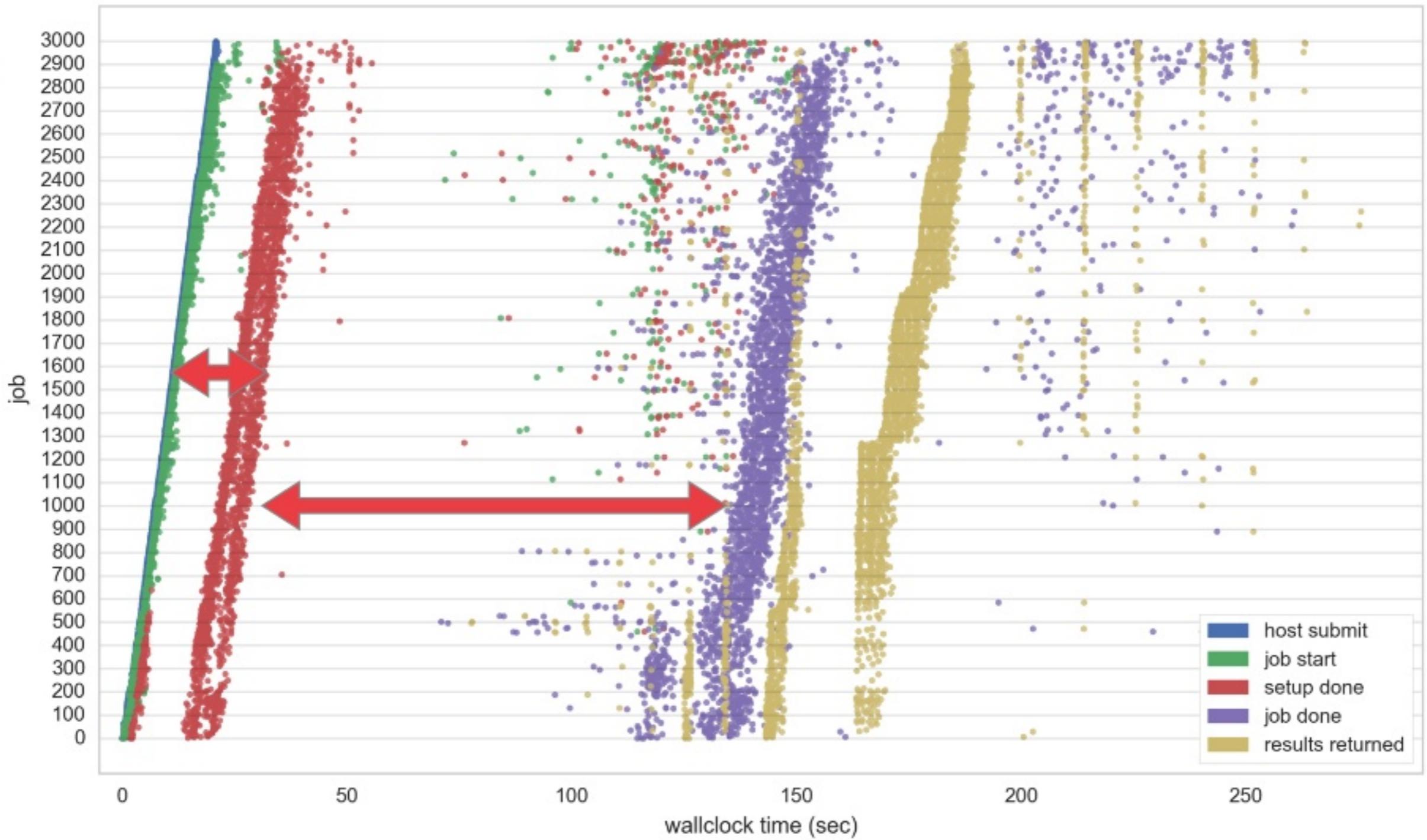


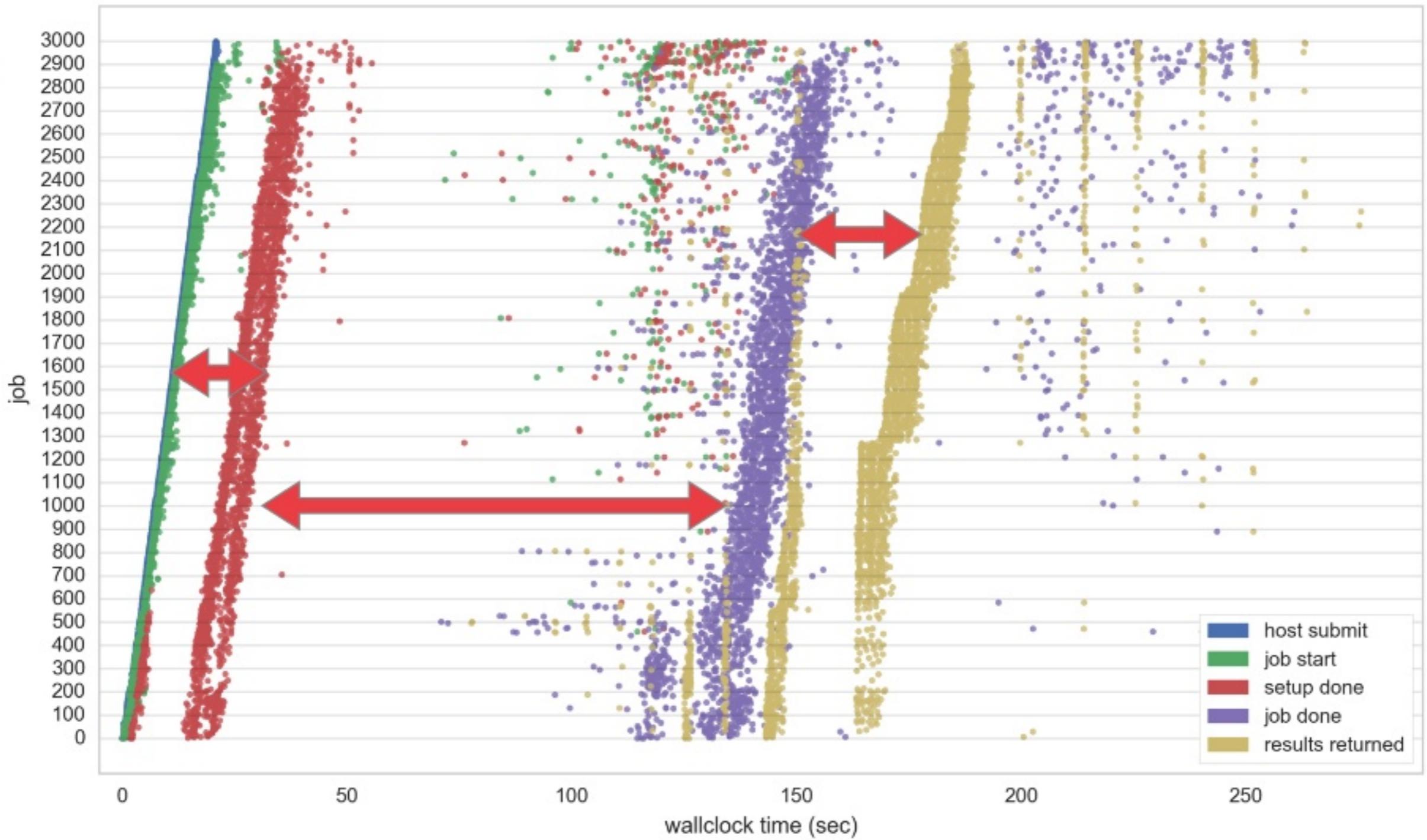
# Preprocess 1.4M images from **IMAGENET**

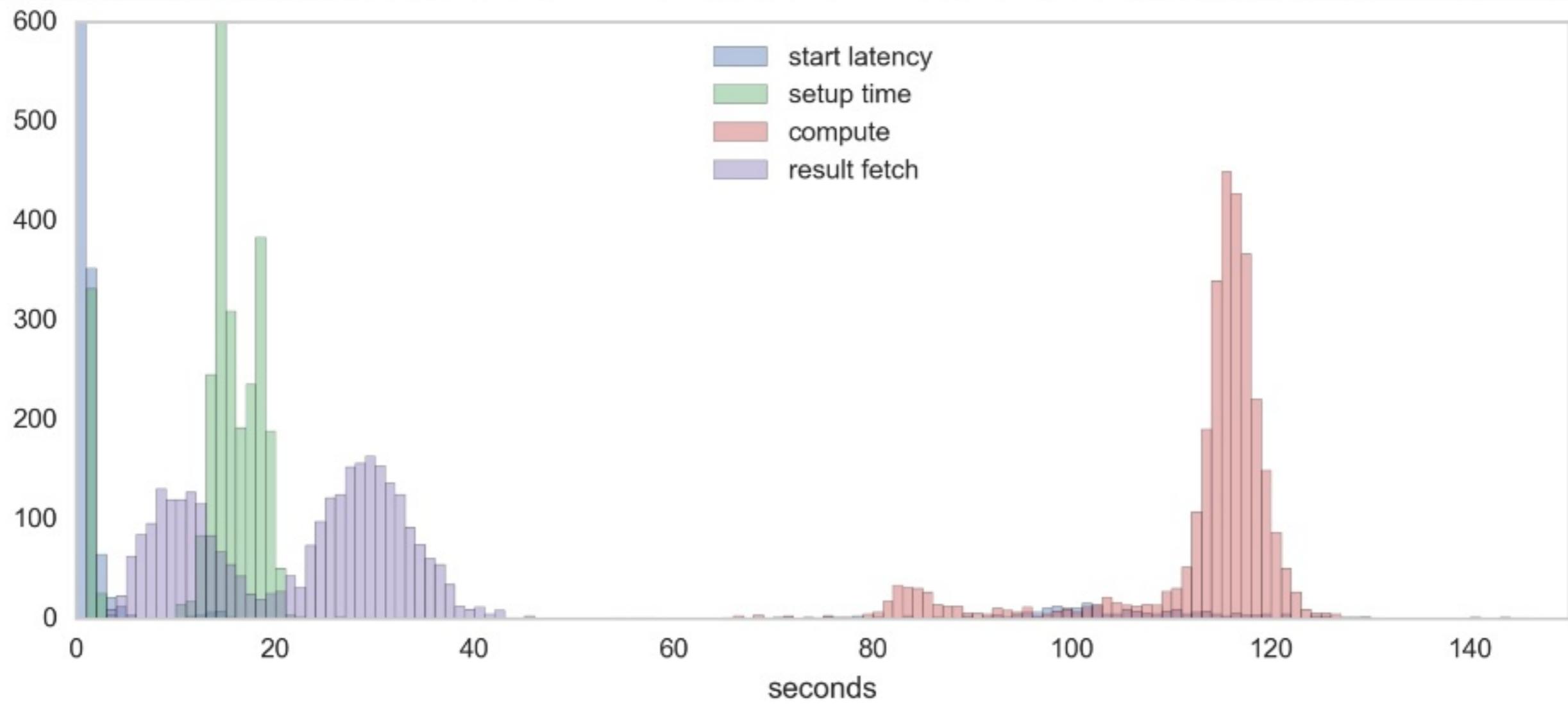
Compute GIST  
image descriptor  
(some random  
python code off  
the internet)

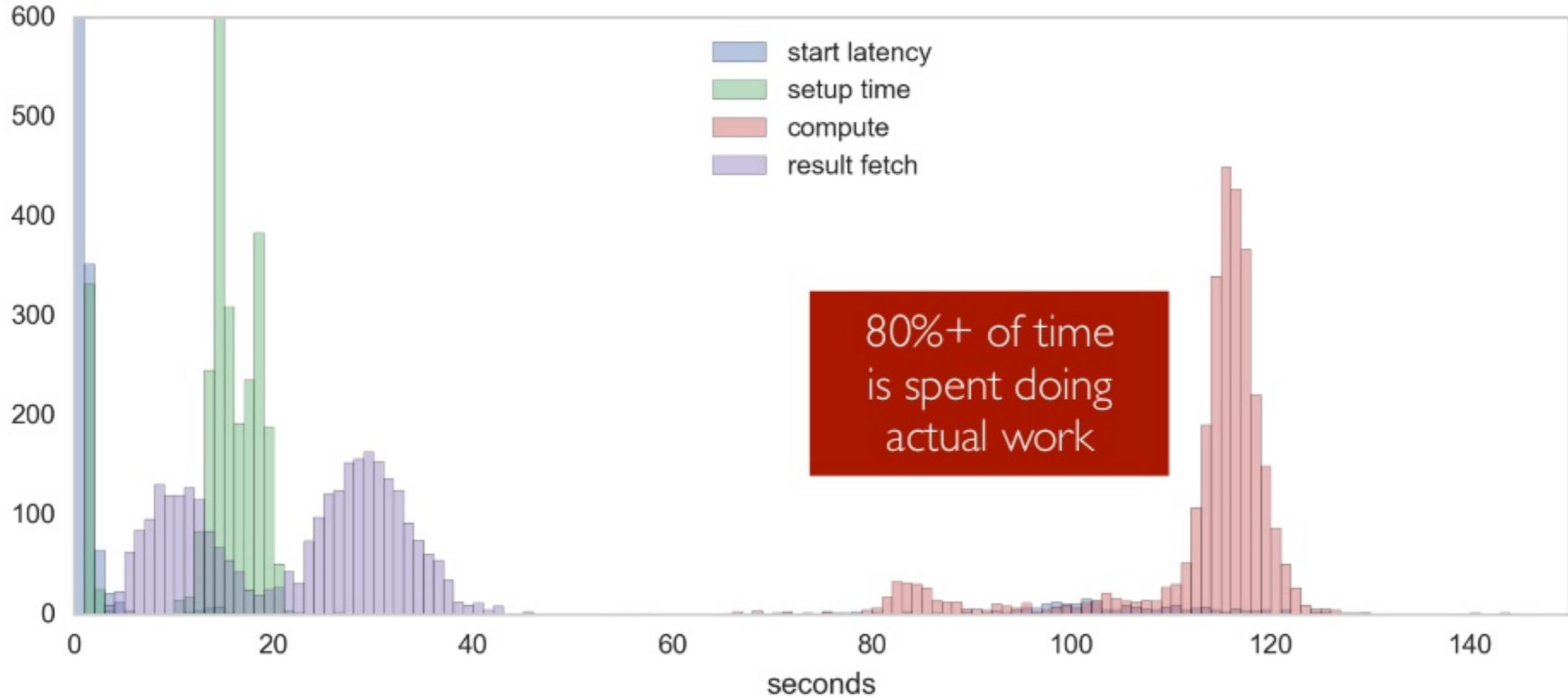












# HOW IT WORKS

your laptop

the cloud

# HOW IT WORKS

```
future = runner.map(fn, data)
```

your laptop

the cloud

# HOW IT WORKS

```
future = runner.map(fn, data)
```

Serialize func and data

your laptop

the cloud

# HOW IT WORKS

```
future = runner.map(fn, data)
```

Serialize func and data

Put on S3

func

data

your laptop

the cloud

# HOW IT WORKS

```
future = runner.map(fn, data)
```

Serialize func and data

Put on S3

Invoke Lambda



your laptop

the cloud

# HOW IT WORKS

```
future = runner.map(fn, data)
```

Serialize func and data

Put on S3

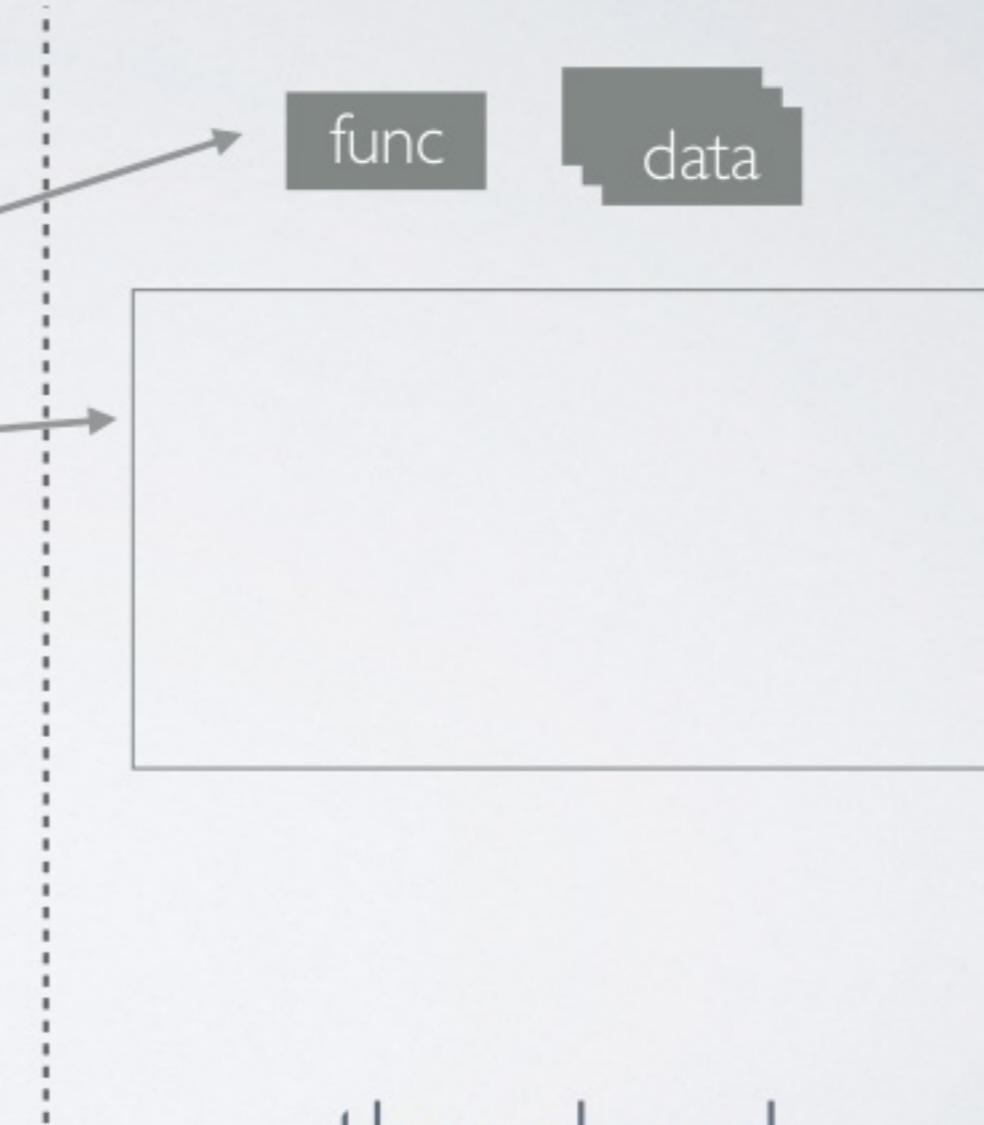
Invoke Lambda

func

data

your laptop

the cloud



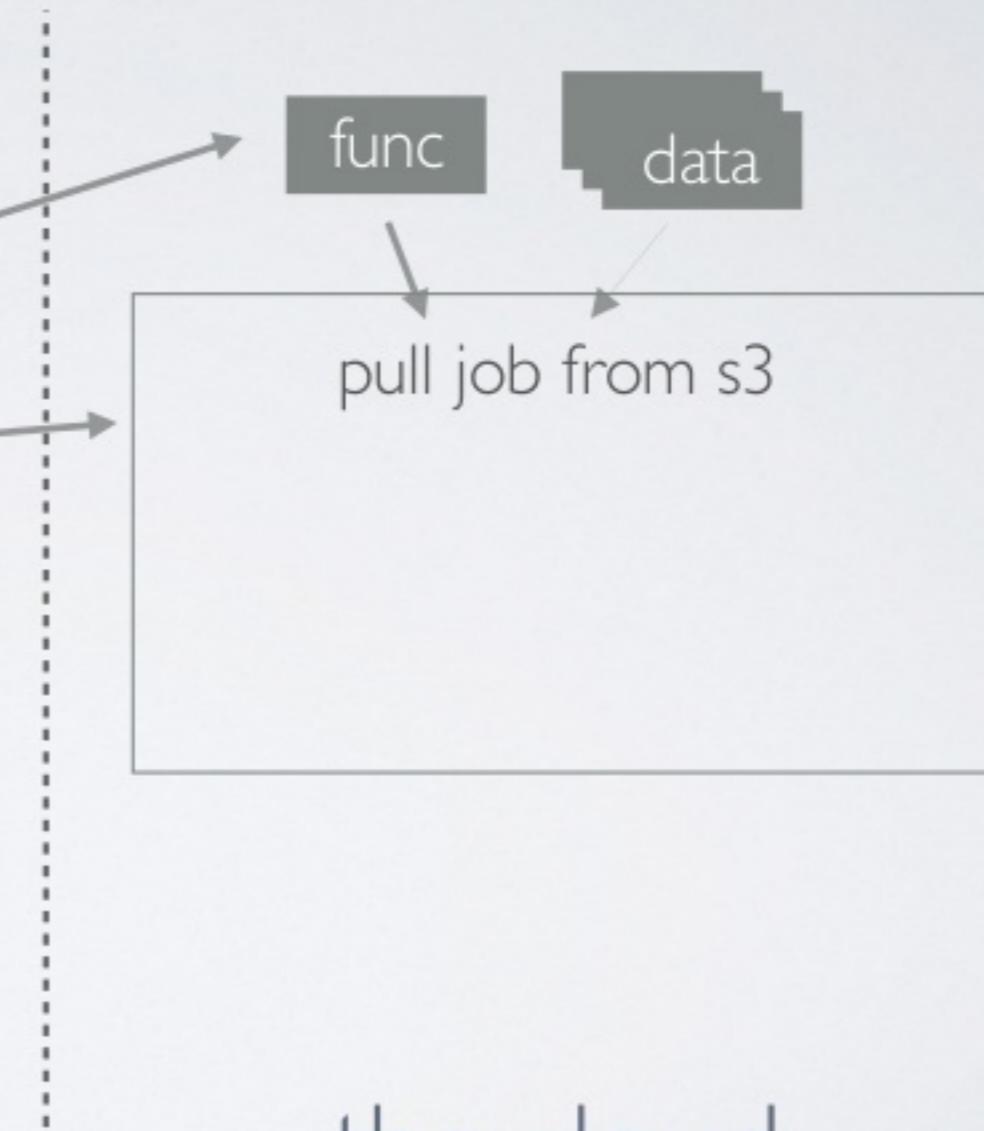
# HOW IT WORKS

```
future = runner.map(fn, data)
```

Serialize func and data

Put on S3

Invoke Lambda



your laptop

the cloud

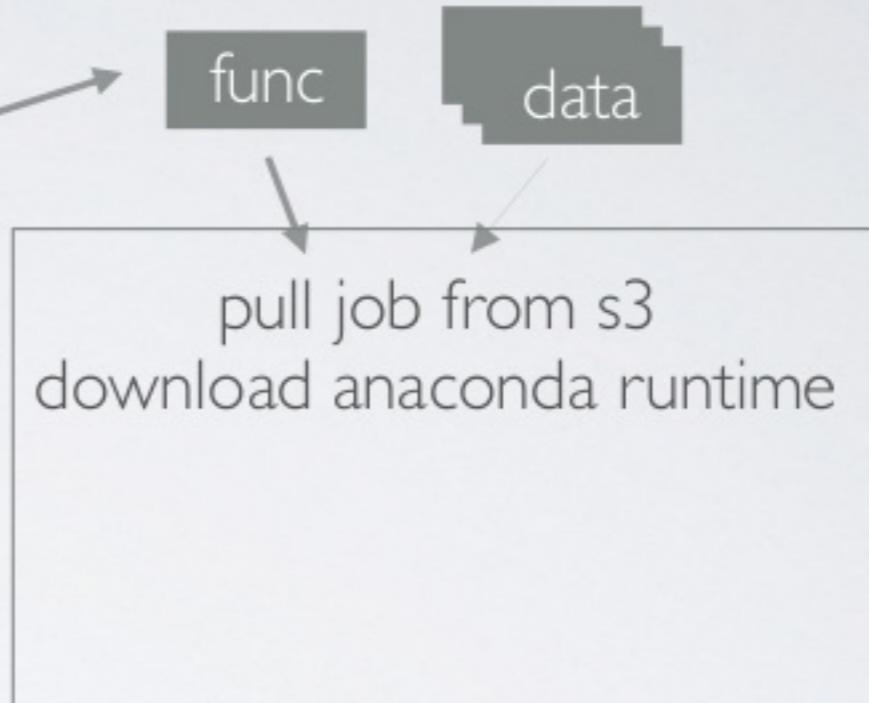
# HOW IT WORKS

```
future = runner.map(fn, data)
```

Serialize func and data

Put on S3

Invoke Lambda



your laptop

the cloud

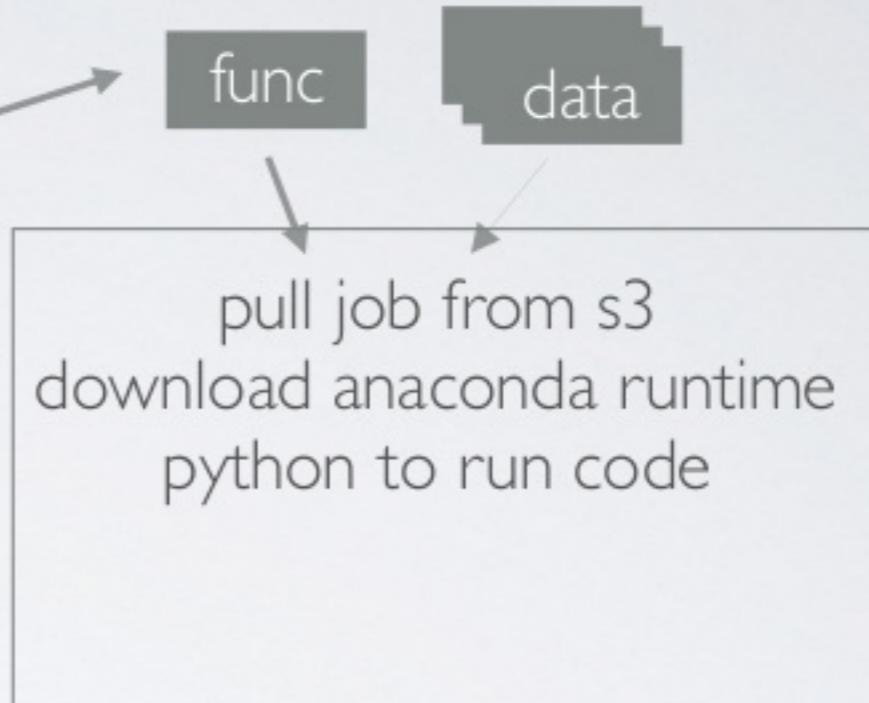
# HOW IT WORKS

```
future = runner.map(fn, data)
```

Serialize func and data

Put on S3

Invoke Lambda



your laptop

the cloud

# HOW IT WORKS

```
future = runner.map(fn, data)
```

Serialize func and data

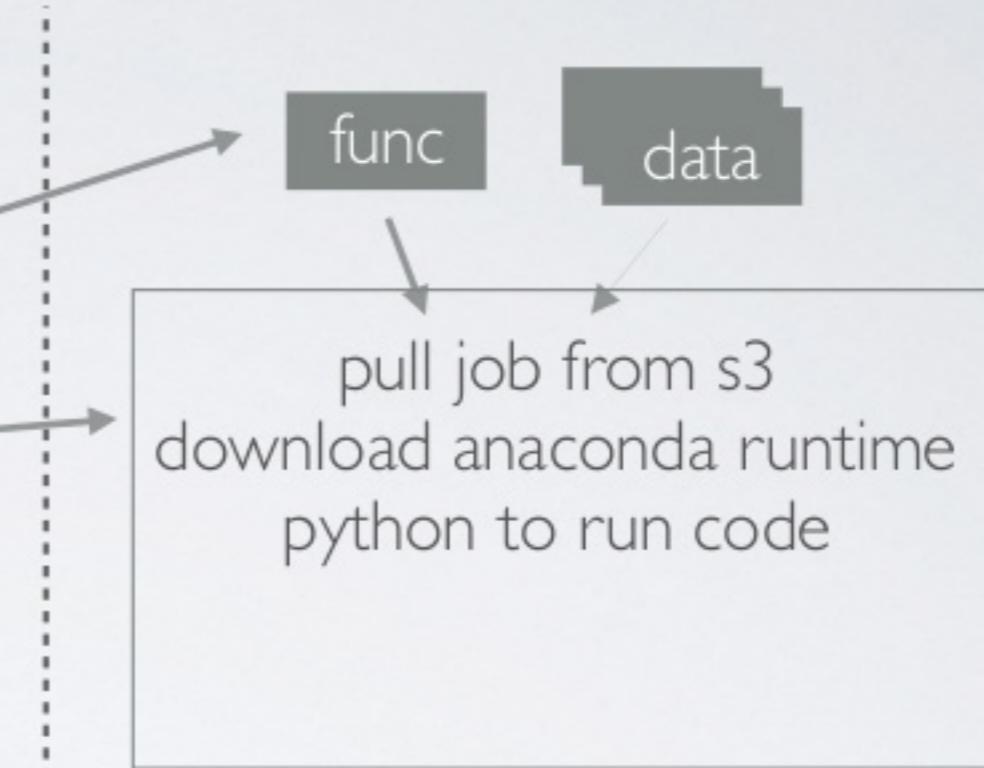
Put on S3

Invoke Lambda

```
future.result()
```

your laptop

the cloud



# HOW IT WORKS

```
future = runner.map(fn, data)
```

Serialize func and data

Put on S3

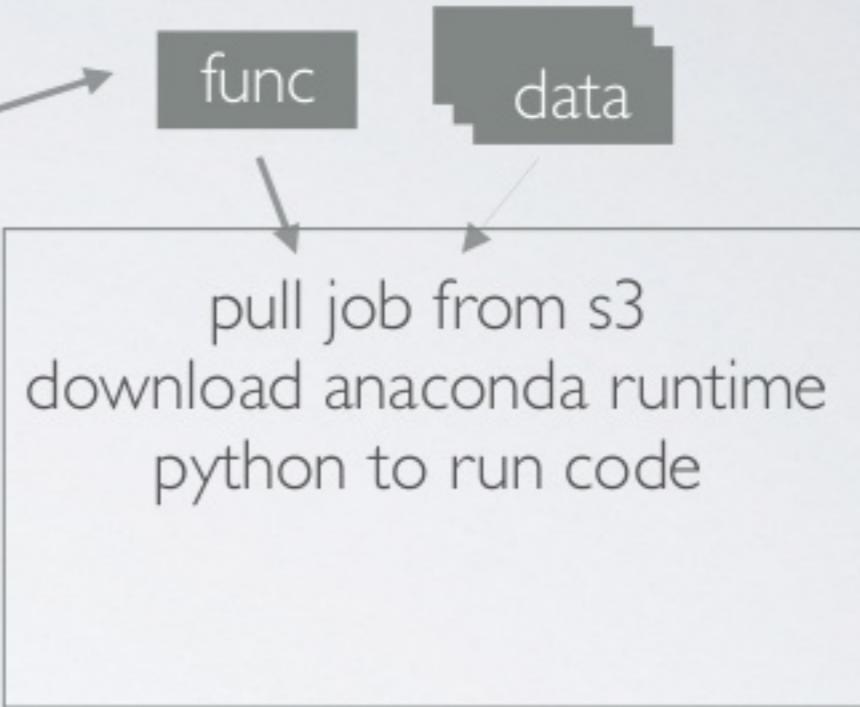
Invoke Lambda

```
future.result()
```

poll S3

your laptop

the cloud



# HOW IT WORKS

```
future = runner.map(fn, data)
```

Serialize func and data

Put on S3

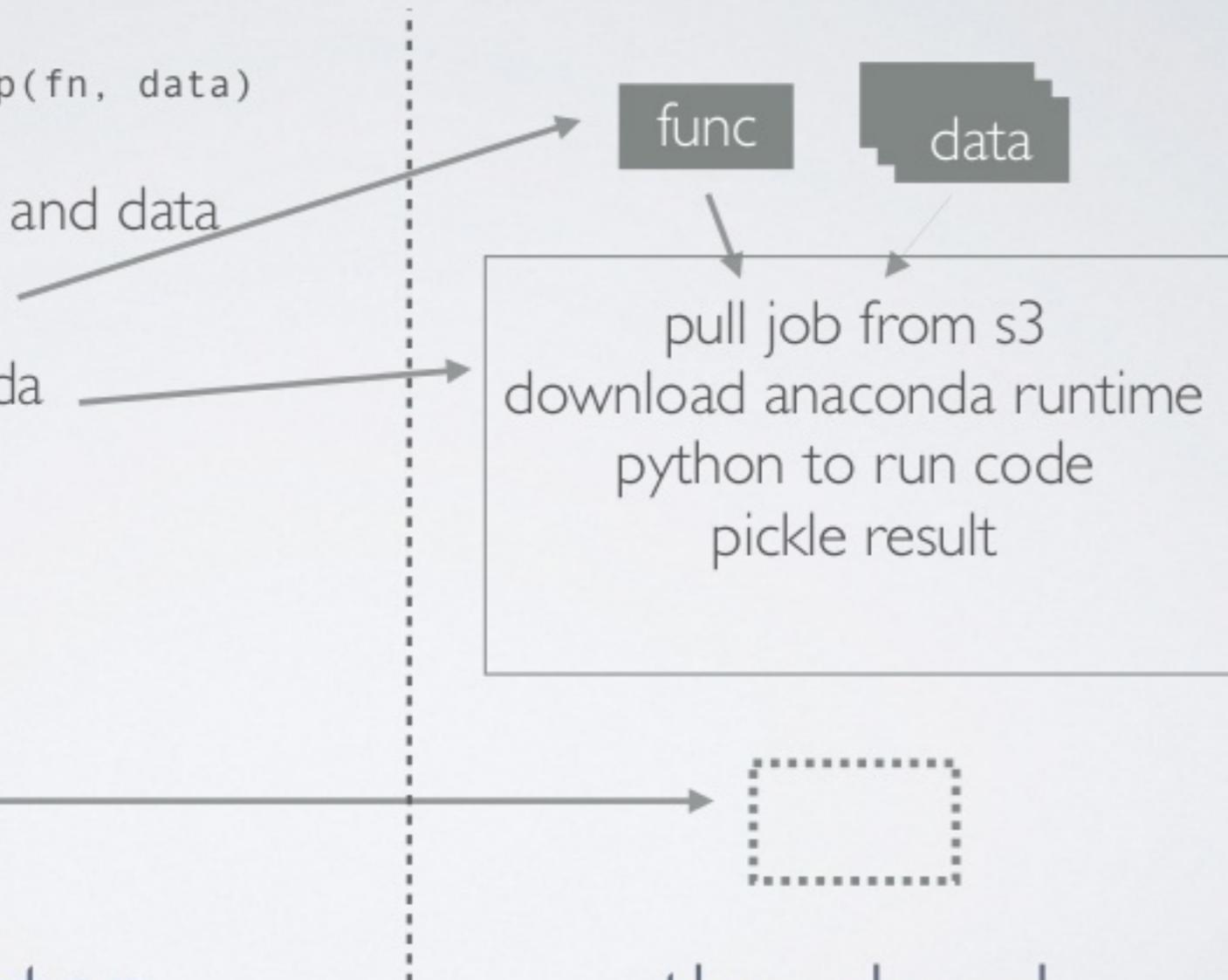
Invoke Lambda

```
future.result()
```

poll S3

your laptop

the cloud



# HOW IT WORKS

```
future = runner.map(fn, data)
```

Serialize func and data

Put on S3

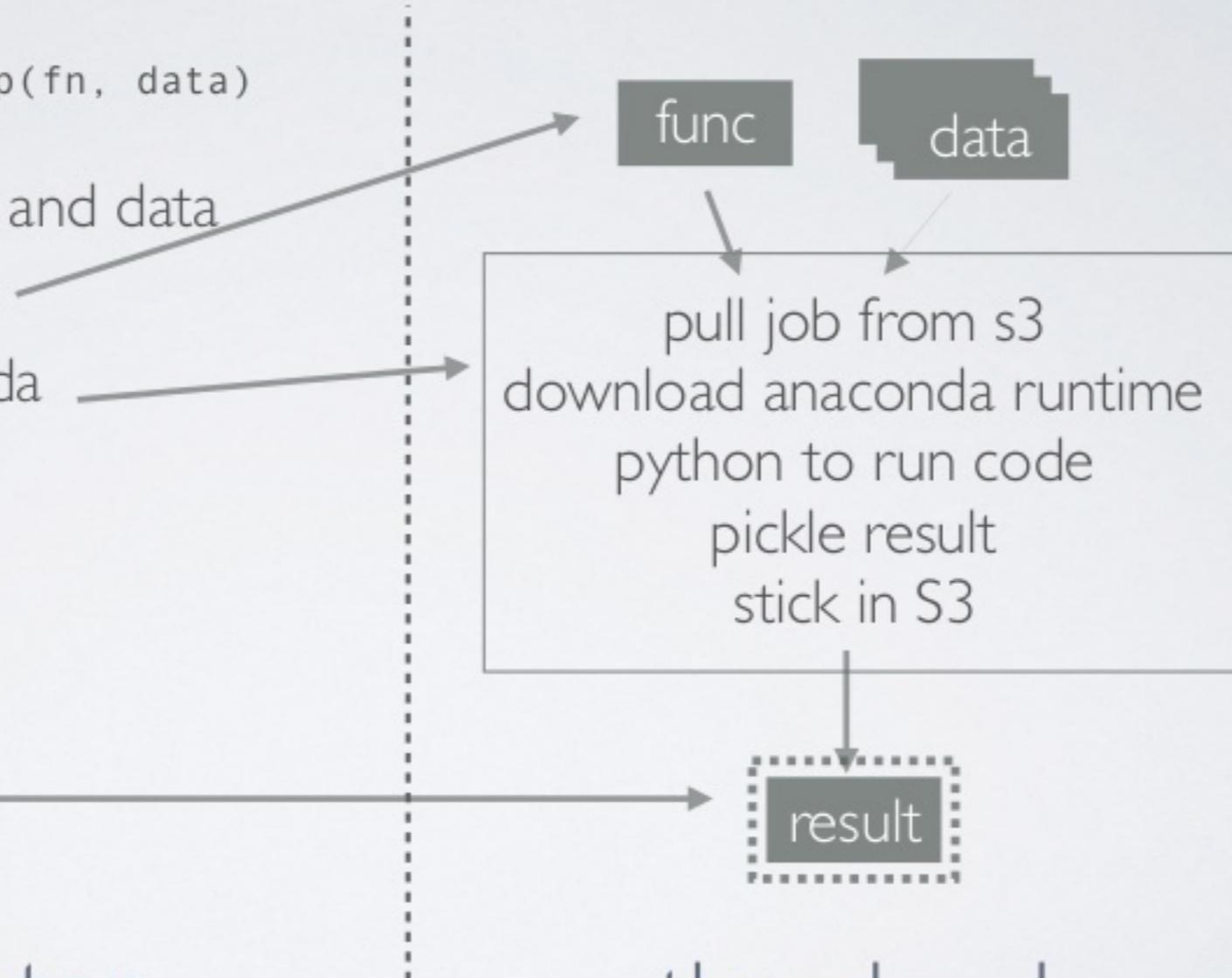
Invoke Lambda

```
future.result()
```

poll S3

your laptop

the cloud



# HOW IT WORKS

```
future = runner.map(fn, data)
```

Serialize func and data

Put on S3

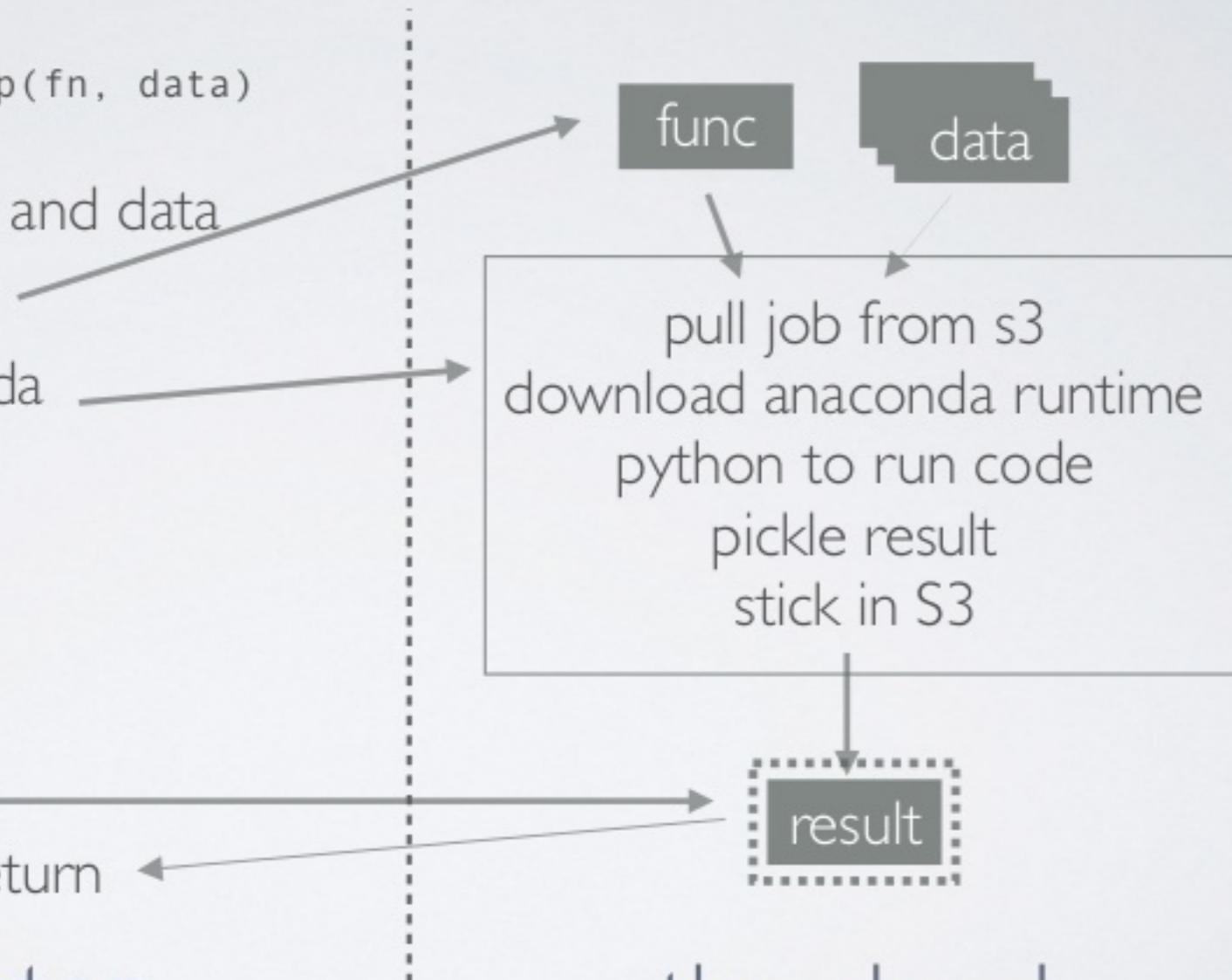
Invoke Lambda

```
future.result()
```

poll S3

unpickle and return

your laptop



the cloud

# THE ROLE OF SERIALIZATION

# THE ROLE OF SERIALIZATION

- Python's dynamism is a help!

# THE ROLE OF SERIALIZATION

- Python's dynamism is a help!
- default serialization method, “`pickle`”, doesn't support everything, leading to `cloudpickle`

# THE ROLE OF SERIALIZATION

- Python's dynamism is a help!
- default serialization method, “`pickle`”, doesn't support everything, leading to `cloudpickle`
- How to handle modules? recursively walk and package

# THE ROLE OF SERIALIZATION

- Python's dynamism is a help!
- default serialization method, “`pickle`”, doesn't support everything, leading to `cloudpickle`
- How to handle modules? recursively walk and package
- C-dependencies don't work, so must build **runtimes**



(*Leptotyphlops carlae*)



(*Leptotyphlops carlcae*)

Want our runtime to include



NumPy



SciPy



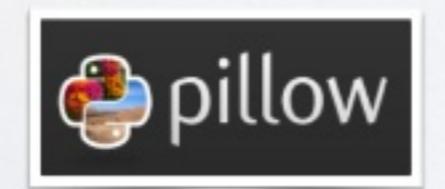
Cython



Numba



scikit  
learn



pillow



(*Leptotyphlops carlcae*)

Start

1205MB

Want our runtime to include



NumPy



SciPy



Cython



Numba



scikit  
learn



pillow



(*Leptotyphlops carlcae*)

Start

1205MB

conda clean

Want our runtime to include



NumPy



SciPy



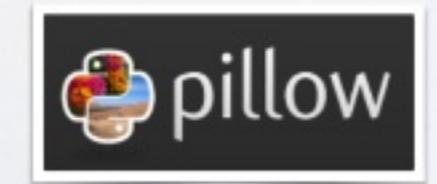
Cython



Numba



scikit  
learn



pillow



(*Leptotyphlops carlcae*)

Start

1205MB

conda clean

977 MB

Want our runtime to include



NumPy



SciPy



Cython



Numba



scikit  
learn



pillow



(*Leptotyphlops carlcae*)

Start

1205MB

conda clean

977 MB

eliminate pkg

Want our runtime to include



NumPy



SciPy



Cython



Numba



scikit  
learn



pillow



(*Leptotyphlops carlcae*)

Start

1205MB

conda clean

977 MB

eliminate pkg

946 MB

Want our runtime to include



NumPy



SciPy



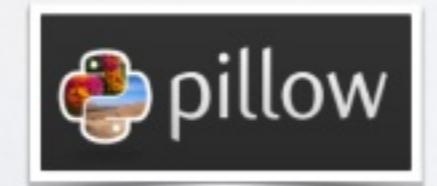
Cython



Numba



scikit  
learn



pillow



(*Leptotyphlops carlae*)

Want our runtime to include



NumPy



SciPy



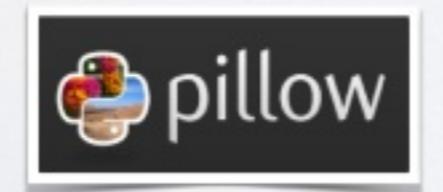
Cython



Numba



scikit  
learn



pillow

Start

1205MB

conda clean

977 MB

eliminate pkg

946 MB

Delete non-AVX2 MKL



(*Leptotyphlops carlcae*)

Want our runtime to include



NumPy



SciPy



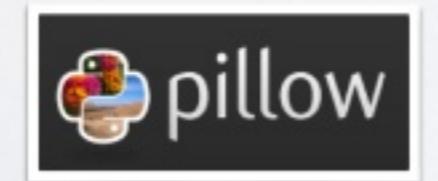
Cython



Numba



scikit  
learn



pillow

Start

1205MB

conda clean

977 MB

eliminate pkg

946 MB

Delete non-AVX2 MKL

670 MB



(*Leptotyphlops carlae*)

Want our runtime to include



NumPy



SciPy



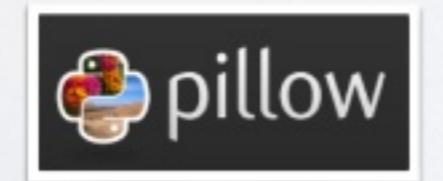
Cython



Numba



scikit  
learn



pillow

Start

1205MB

conda clean

977 MB

eliminate pkg

946 MB

Delete non-AVX2 MKL

670 MB

strip shared libs



(*Leptotyphlops carlcae*)

Want our runtime to include



NumPy



SciPy



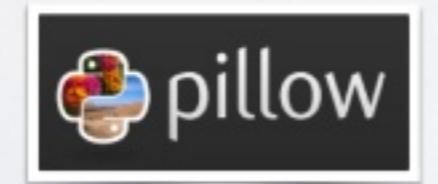
Cython



Numba



scikit  
learn



pillow

Start

1205MB

conda clean

977 MB

eliminate pkg

946 MB

Delete non-AVX2 MKL

670 MB

strip shared libs

510MB



(*Leptotyphlops carlcae*)

Want our runtime to include



NumPy



SciPy



Cython



Numba



scikit  
learn



pillow

Start

1205MB

conda clean

977 MB

eliminate pkg

946 MB

Delete non-AVX2 MKL

670 MB

strip shared libs

510MB

delete pyc



(*Leptotyphlops carlae*)

Want our runtime to include



NumPy



SciPy



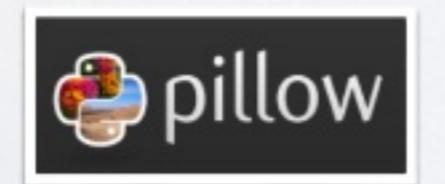
Cython



Numba



scikit  
learn



pillow

Start

1205MB

conda clean

977 MB

eliminate pkg

946 MB

Delete non-AVX2 MKL

670 MB

strip shared libs

510MB

delete pyc

441MB

# LIMITATIONS (AND IDEAS FOR SOLUTIONS)

- Lambda limits
- S3 Limits
- richer operations

# LIMITATIONS (AND IDEAS FOR SOLUTIONS)

- Lambda limits
- S3 Limits
- **richer operations**

MAP IS NOT ENOUGH

# MAP IS NOT ENOUGH?

A lot of data analytics looks like:

# MAP IS NOT ENOUGH?

A lot of data analytics looks like:

Data

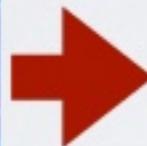


# MAP IS NOT ENOUGH?

A lot of data analytics looks like:

Data

ETL /  
preprocessing



# MAP IS NOT ENOUGH?

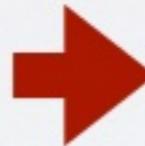
A lot of data analytics looks like:

Data



ETL /  
preprocessing

featurization



# MAP IS NOT ENOUGH?

A lot of data analytics looks like:

Data



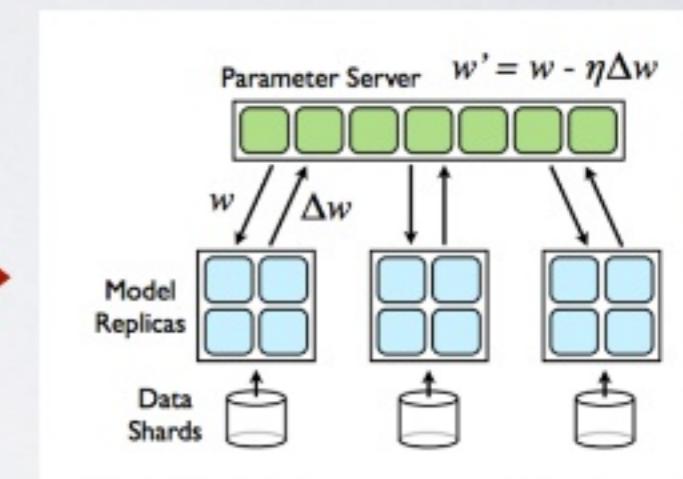
ETL /  
preprocessing



featurization



machine learning



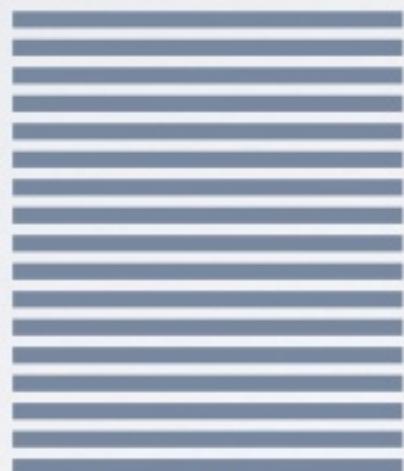
# MAP IS NOT ENOUGH?

A lot of data analytics looks like:

Data



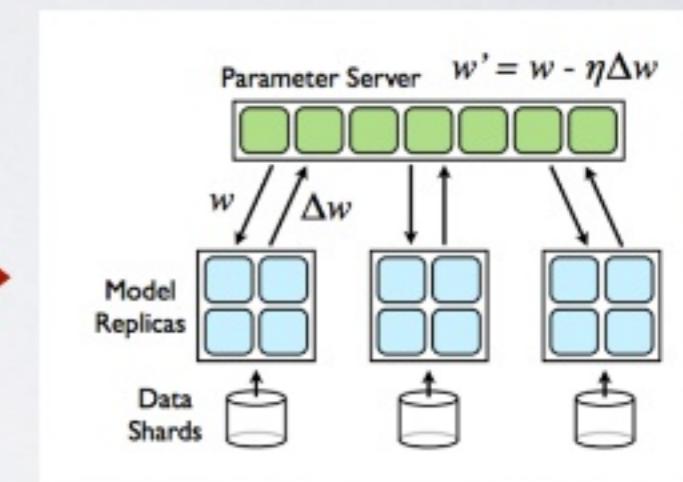
ETL /  
preprocessing



featurization



machine learning



Great PyWren Fit

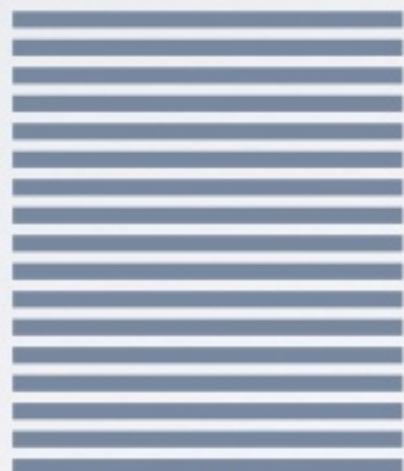
# MAP IS NOT ENOUGH?

A lot of data analytics looks like:

Data



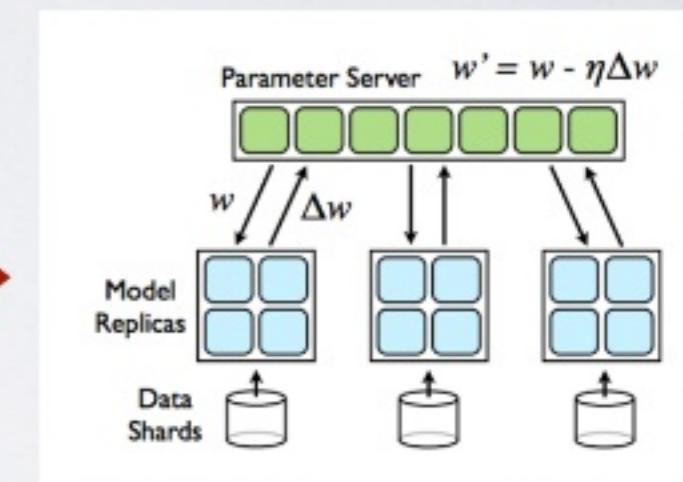
ETL /  
preprocessing



featurization



machine learning



Great PyWren Fit

Distributed!  
Scale! TensorFlow  
Deep MLBase

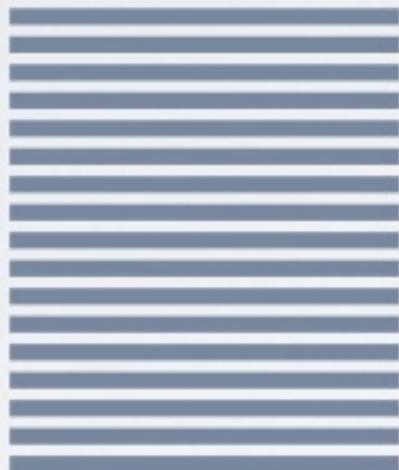
# MAP IS NOT ENOUGH?

A lot of data analytics looks like:

Data



ETL /  
preprocessing



featurization



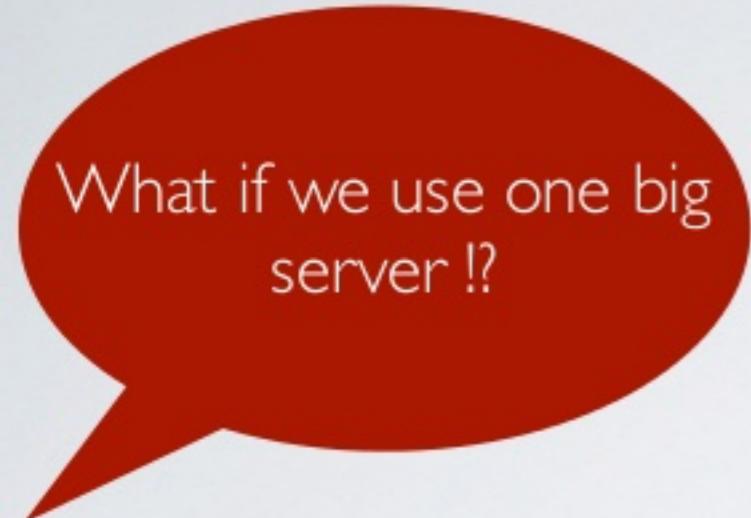
machine learning



Great PyWren Fit

Distributed!  
Scale! TensorFlow  
Deep MLBase

# SINGLE-MACHINE REDUCE



What if we use one big server !?

# SINGLE-MACHINE REDUCE



|             | cores           | RAM    | COST    |
|-------------|-----------------|--------|---------|
| x1.32xlarge | 64              | 2 TB   | \$14/hr |
| x1.16xlarge | 32              | 1 TB   | \$7/hr  |
| p2.16xlarge | 32 +<br>16 GPUs | 750 GB | \$14/hr |
| r4.16xlarge | 32              | 500 GB | \$4/hr  |

# SINGLE-MACHINE REDUCE



|             | cores           | RAM    | COST    |
|-------------|-----------------|--------|---------|
| x1.32xlarge | 64              | 2 TB   | \$14/hr |
| x1.16xlarge | 32              | 1 TB   | \$7/hr  |
| p2.16xlarge | 32 +<br>16 GPUs | 750 GB | \$14/hr |
| r4.16xlarge | 32              | 500 GB | \$4/hr  |

```
futures = exec.map(function, data)
```

```
answer = exec.reduce(reduce_func, futures)
```

WHAT ABOUT SHUFFLES ?

# Sort Benchmark Home Page

**New:** The 2016 records are listed below in green. Thank you to all of the many 2016 entrants!

## Background

Until 2007, the sort benchmarks were primarily defined, sponsored and administered by Jim Gray. Following Jim's disappearance at sea in January 2007, the sort benchmarks have been continued by a committee of past colleagues and sort benchmark winners. The Sort Benchmark committee members include:

- Chris Nyberg of Ordinal Technology Corp
- Mehul Shah of Amazon Web Services
- Naga Govindaraju of Microsoft

## Top Results

|       | Daytona   | Indy   |
|-------|---|--|
| Gray  | 2016, 44.8 TB/min<br><br><b>Tencent Sort</b><br>100 TB in 134 Seconds<br>512 nodes x (2 OpenPOWER 10-core POWER8 2.926 GHz,<br>512 GB memory, 4x Huawei ES3600P V3 1.2TB NVMe SSD,<br>100Gb Mellanox ConnectX4-EN)<br>Jie Jiang, Lixiong Zheng, Junfeng Pu,<br>Xiong Cheng, Chongqing Zhao<br>Tencent Corporation<br>Mark R. Nutter, Jeremy D. Schaub | 2016, 60.7 TB/min<br><br><b>Tencent Sort</b><br>100 TB in 98.8 Seconds<br>512 nodes x (2 OpenPOWER 10-core POWER8 2.926 GHz,<br>512 GB memory, 4x Huawei ES3600P V3 1.2TB NVMe SSD,<br>100Gb Mellanox ConnectX4-EN)<br>Jie Jiang, Lixiong Zheng, Junfeng Pu,<br>Xiong Cheng, Chongqing Zhao<br>Tencent Corporation<br>Mark R. Nutter, Jeremy D. Schaub |
| Cloud | 2016, \$1.44 / TB<br><br><b>NADSort</b><br>100 TB for \$144<br>394 Alibaba Cloud ECS ecs.n1.large nodes x<br>(Haswell E5-2680 v3, 8 GB memory,<br>40GB Ultra Cloud Disk, 4x 135GB SSD Cloud Disk)<br>Qian Wang, Rong Gu, Yihua Huang<br>Nanjing University<br>Reynold Xin<br>Databricks Inc.<br>Wei Wu, Jun Song, Junluan Xia<br>Alibaba Group Inc.   | 2016, \$1.44 / TB<br><br><b>NADSort</b><br>100 TB for \$144<br>394 Alibaba Cloud ECS ecs.n1.large nodes x<br>(Haswell E5-2680 v3, 8 GB memory,<br>40GB Ultra Cloud Disk, 4x 135GB SSD Cloud Disk)<br>Qian Wang, Rong Gu, Yihua Huang<br>Nanjing University<br>Reynold Xin<br>Databricks Inc.<br>Wei Wu, Jun Song, Junluan Xia<br>Alibaba Group Inc.    |

Gray

2014, 4.35 TB/min

## TritonSort

100 TB in 1,378 seconds

186 Amazon EC2 i2.8xlarge nodes x

(32 vCores - 2.50Ghz Intel Xeon E5-2670 v2, 244GB memory, 8x800 GB SSD)

[Michael Conley](#), [Amin Vahdat](#),  
[George Porter](#)

University of California, San Diego

2014, 4.27 TB/min

## Apache Spark

100 TB in 1,406 seconds

207 Amazon EC2 i2.8xlarge nodes x

(32 vCores - 2.5Ghz Intel Xeon E5-2670 v2, 244GB memory, 8x800 GB SSD)

Reynold Xin, Parviz Deyhim, Xiangrui Meng,  
Ali Ghodsi, Matei Zaharia

Databricks

2013, 1.42 TB/min

## Hadoop

102.5 TB in 4,328 seconds

2100 nodes x

(2 2.3Ghz hexcore Xeon E5-2630, 64 GB memory, 12x3TB disks)

[Thomas Graves](#)  
Yahoo! Inc.

2011, 0.725 TB/min

## TritonSort

100 TB in 8,274 seconds

52 nodes x

# TERASORT IN PYWREN

S3

# TERASORT IN PYWREN

S3

**map**(partition\_keys)

# TERASORT IN PYWREN

S3

**map**(partition\_keys)

Elastic Cache  
(Redis)

# TERASORT IN PYWREN

S3

**map**(partition\_keys)

Elastic Cache  
(Redis)

**map**(sort\_partition)

# TERASORT IN PYWREN

S3

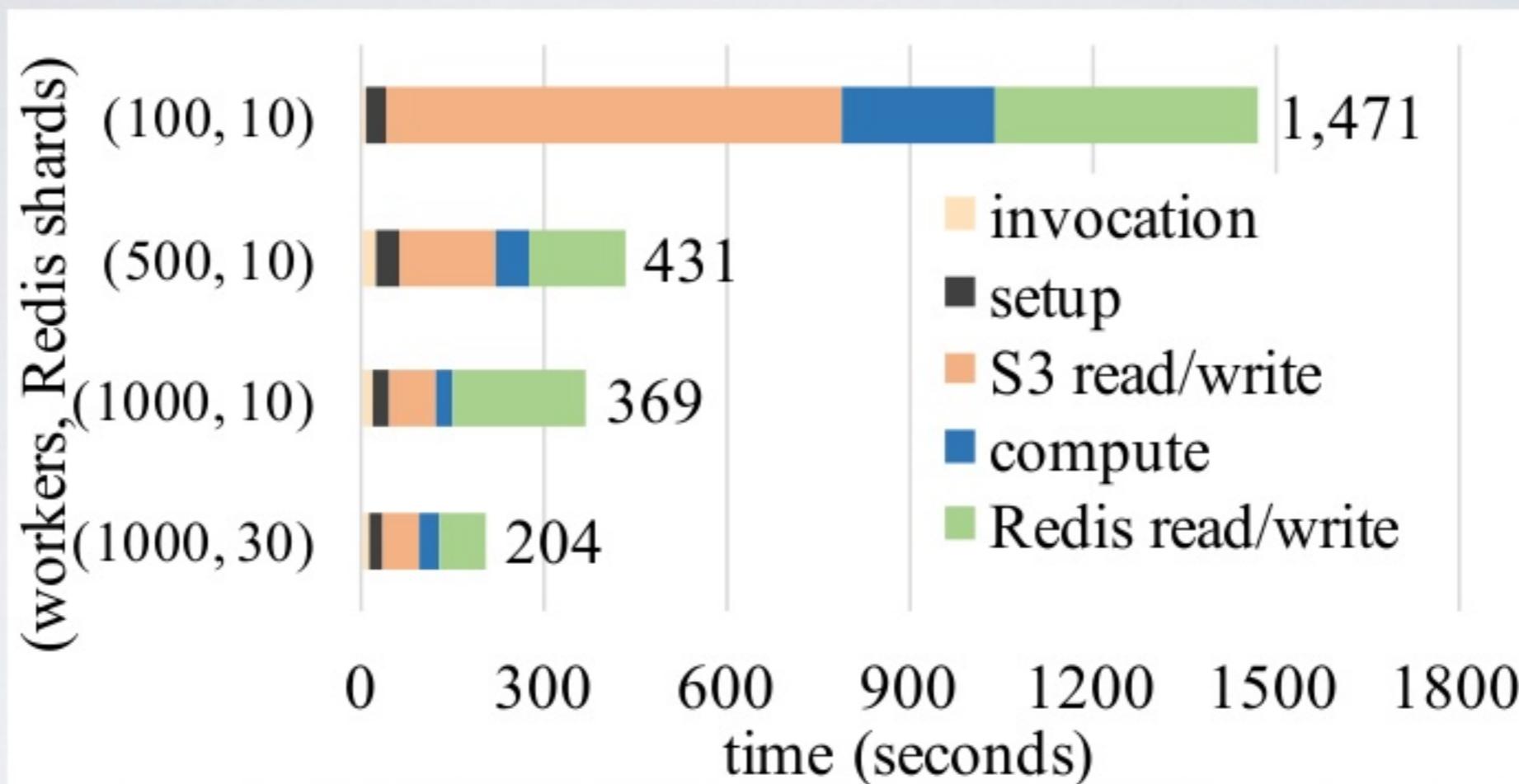
**map**(partition\_keys)

Elastic Cache  
(Redis)

**map**(sort\_partition)

S3

# I TB SORT



# WIP: SCHEDULING ENGINE

# WIP: SCHEDULING ENGINE

- Rate-limiting of invocation
- intelligent retry behavior (user-specified)
- straggler detection and mitigation

# THE SERVERLESS FUTURE

# THE SERVERLESS FUTURE

Where we are going:

# THE SERVERLESS FUTURE

- Users Users Users (You !!!)

Where we are going:

# THE SERVERLESS FUTURE

- Users Users Users (You !!!)
- Circumventing limitations

Where we are going:

# THE SERVERLESS FUTURE

Where we are going:

- Users Users Users (You !!!)
- Circumventing limitations
- Make serialization rock-solid

# THE SERVERLESS FUTURE

Where we are going:

- Users Users Users (You !!!)
- Circumventing limitations
- Make serialization rock-solid
- Richer primitives!

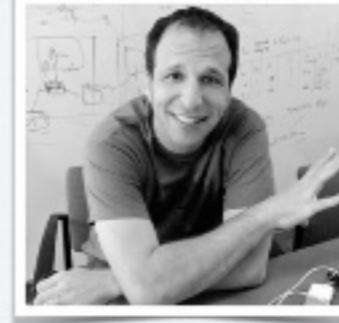
# Effortlessly scaling data science with **PyWren**



Eric Jonas



Qifan Pu



Ben Recht



Ion Stoica

# Effortlessly scaling data science with **PyWren**

<http://pywren.io>

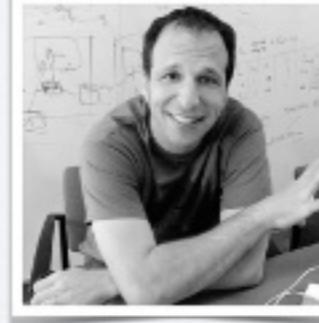
<https://github.com/pywren/pywren>



Eric Jonas



Qifan Pu



Ben Recht



Ion Stoica