

Devices-as-Services and The New Internet as a Platform of Things

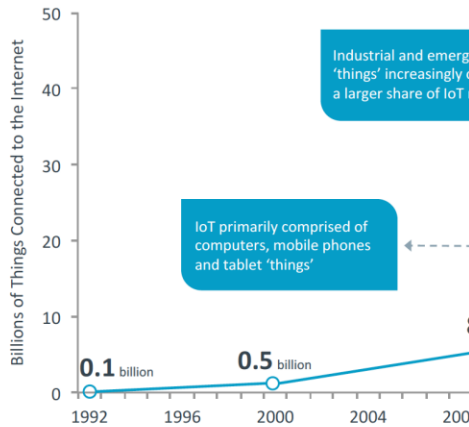


Rich Wolski
Computer Science Department
University of California, Santa Barbara
September 17, 2019

Internet of Things

Projecting the 'Things' Behind the Internet of Things

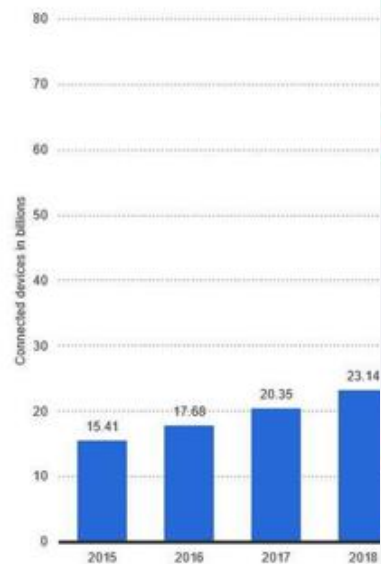
From 2014-2020, IoT grows at an annual compound rate of 23.1% CAGR



CompTIA

Internet of Things - number of connected devices

Internet of Things (IoT) connected devices 2015 to 2025 (in billions)



1 trillion

objects expected to connect to the Internet by 2025

ADOPTION OF NEW TECHNOLOGIES IS ACCELERATING

Time to reach 50 million users, years

Technology	Time to reach 50 million users (years)
Radio	36
TV	13
iPod	4
Internet	3
Facebook	1
Twitter	.75

TECHNOLOGICAL BREAKTHROUGHS ARE SPEEDING UP

The path toward mobile internet

Event	Year	Time to reach 50 million users (years)
1876 first phone call	1876	115
1991 first website	1991	16
2007 first iPhone	2007	-

DISRUPTIVE TECHNOLOGIES TO WATCH

Estimated potential economic impact by 2025 across sized applications*

Technology	Estimated potential economic impact by 2025
Mobile Internet	\$4 trillion-\$11 trillion
Automation of knowledge work	\$5 trillion-\$7 trillion
Internet of Things	\$3 trillion-\$6 trillion

*Including consumer surplus; McKinsey Global Institute, Disruptive Technologies, May 2013

MCKINSEY GLOBAL INSTITUTE

Dubin, Margolis, and Woelzel, No Ordinary Disruption: The Four Global Forces Breaking All the Trends, May 2013
Including consumer surplus; McKinsey Global Institute, Disruptive Technologies, May 2013

Three Observations About Putting Things on The Internet

- Power is everything
- Deployment is “soft”-ware
- Build “repurpose” built instead of built.

The Datacenter



FreshDatavention

2018-03-19 16:08:33



IPC

Power is Everything in this Setting

- Power is more expensive to provision than compute, communication, storage
 - For example: ~\$20K to put solar in for 4 PC-class edge systems and long-range WiFi
- Power infrastructure is fragile
 - Dust, sun angle, industrial accidents
- Power infrastructure is invasive
 - Can't trench conduit through the space
 - Solar towers interfere with farming operations
- Scaling properties are terrible
 - More power means more **expensive**, **fragile**, and **invasive**

What's Wrong with this Picture?



The Device is in the Wrong Place

- The sensors were designed for deployment in rocky soil, with all-day sun, at altitude
- \$250K in equipment, 2 person days of installation
- Two separate contracts
 - One for the equipment and software development
 - One for the “siting” and deployment
- Abandoned 18 months after installation

What Went Wrong?

- The software development was considered completely separate from how and where the device was to be deployed.
- Agile development for IoT must include deployment
 - Where the device goes influences the software which influences where the device goes which...
 - *The Deployment Manager missed the scrum*
- For IoT, **Deployment is Software**

What Went Right?

- They donated the equipment to our research group (Thanks!)
 - Built for long-term deployment (still up and running)
- Needs to be reprogrammed for a different purpose
 - Current sensor readings are useless
- Repurposing the device is hard
 - Three data loggers
 - Primitive networking
 - Totally insecure
- **Build “repurpose” built instead of purpose built.**

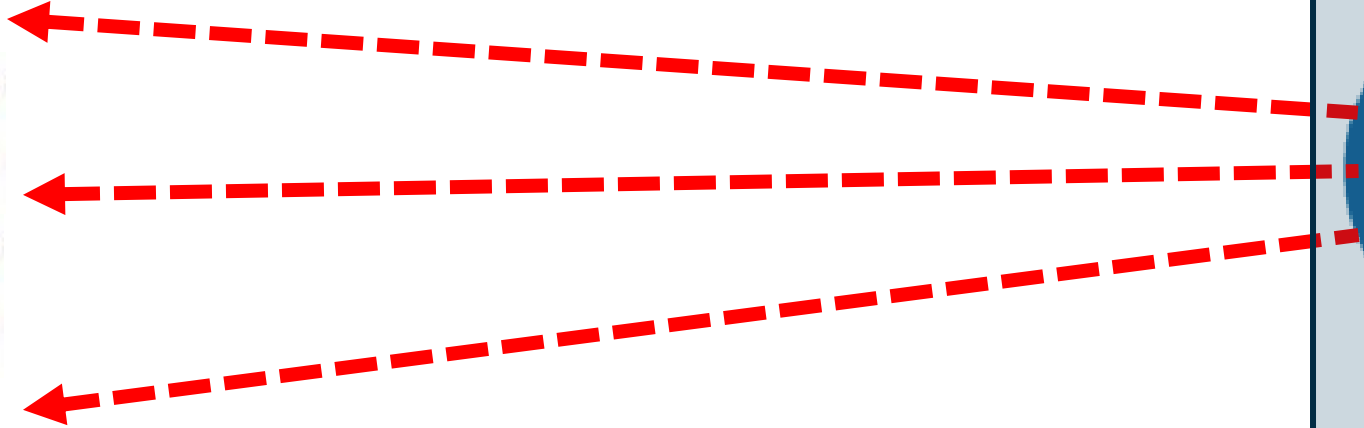
The New World (circa 2025)

- Power is everything
- Deployment is software
- Build repurpose built instead of purpose built
- Conclusion:

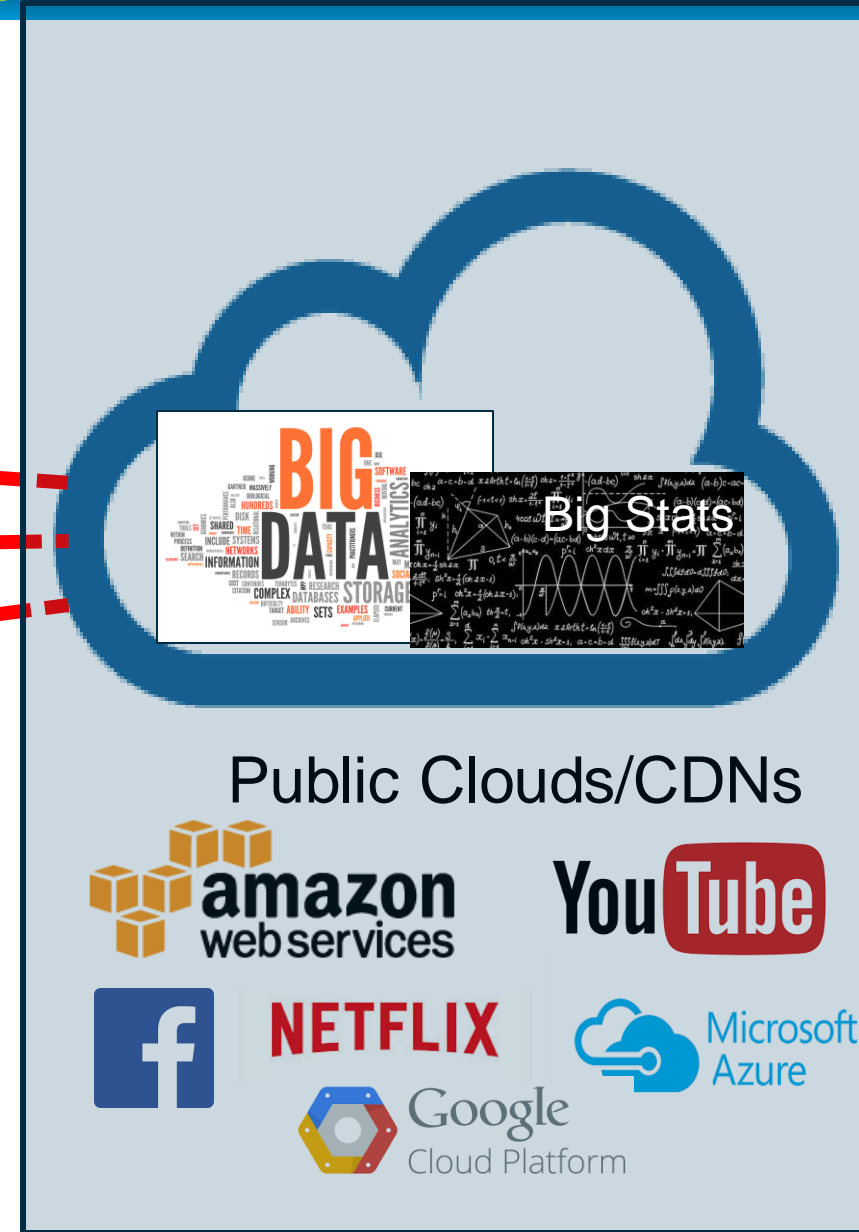
The Internet, at present, is engineered “backwards” from what is required for IoT

Cloud Today = E-Commerce & Entertainment:

Data Moves From Cloud to User Devices



➡ Download (read) dominated:
streaming entertainment/content

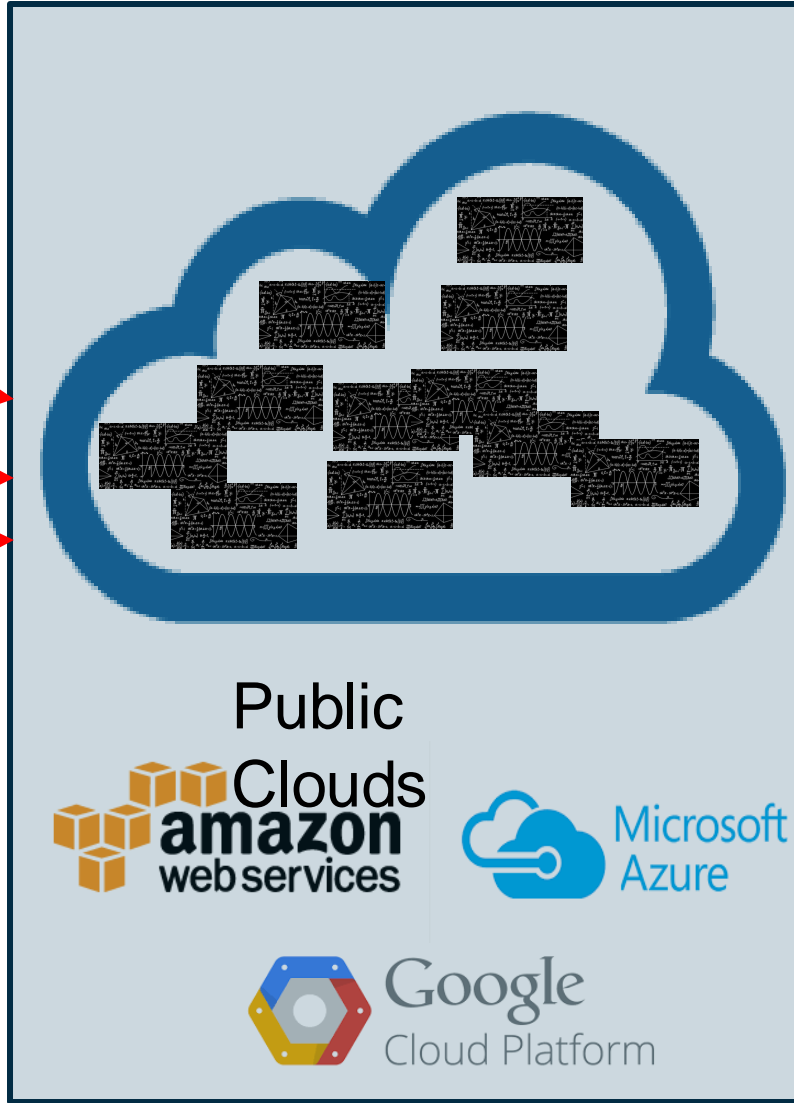
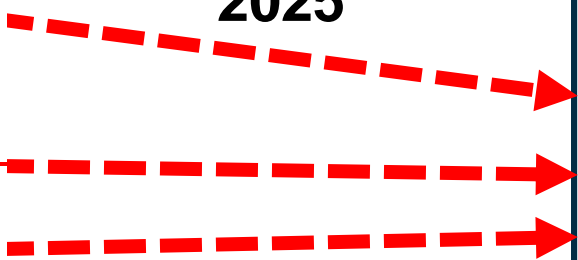


Cloud Today

- Serving consumer content is “read dominated”
 - The “writes” are web page clicks or, at worst, single uploads
- Aggregating click activity generates “Big Data”
- Inference and prediction analytics (Big Stats) “work” because the sample sizes are huge
- Goal: ***Make an inference or a prediction about a specific individual***
 - Big Data and Big Stats are needed because people are so diverse



**'N' Billion
IoT
Devices by
2025**

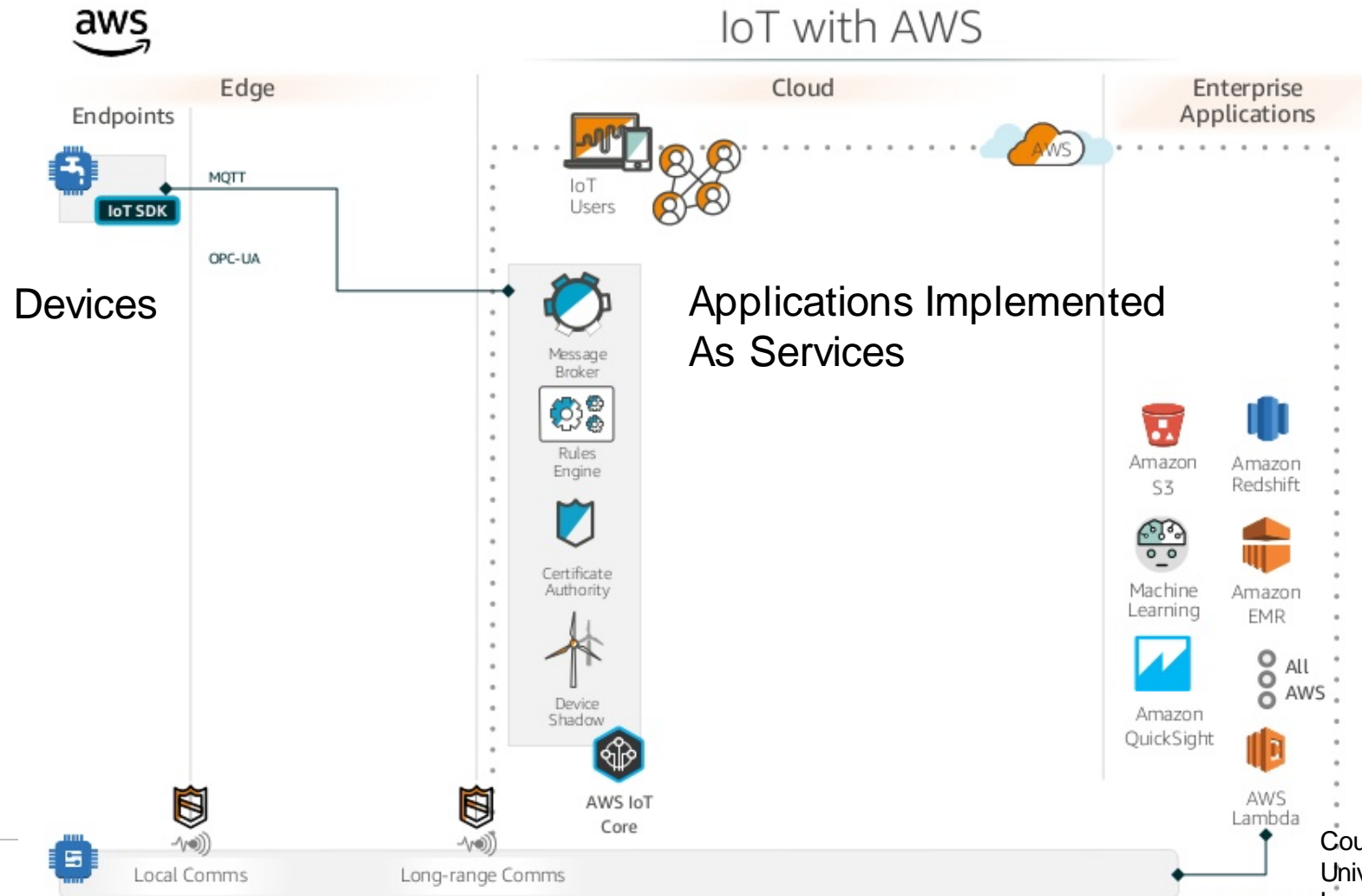


➡ Upload (write)
dominated:
sensing and
monitoring

Not Big Data but Big Little Data

- Data sets and stats are not that big
- **Rich's IoT Data Conjecture**
 - *The actionable relevance of IoT data decays like the square of the distance from where it is gathered*
- Not Big Data but lots and lots of Little Data
- Data is often read only once (to create an aggregate) which makes the data even smaller.
 - Why move the data half way across the country to use it in an aggregate and throw it away?
- **Rich's Data Motto**
 - *The more the data moves, the more expensive, power-using, and failure prone the system is.*

Today: Devices are Cloud Users and Applications are Services



Courtesy Dr. Markus Mock,
University of Applied Sciences,
Landshut, Germany



Why are Cloud Applications Built as Services?

- Services react to user activity
 - Run only when needed => **efficient**
- Services hide the gory detail
 - Cloud comprise hundreds of APIs that are hidden by a service => **programmable deployment**
- Services are reprogrammable
 - Extensions or code modifications happen behind the service interface => **repurposable**

Ergo, for IoT, Devices Should Run Services

- Power efficient: services only respond when they are prompted
- Programmable Deployment: cloud services often use deployment software
- Repurposeable: Devices are multi-purpose and fully programmable
- Idea: **Flip the Internet**
 - Devices run services
 - Edge runs services
 - Cloud runs applications as clients and user interfaces

Devices-as-Services

- Need a service development, deployment, and hosting environment that is **multi-scale**
 - Deployment support: same service (unmodified) can run on device, or edge, or cloud
 - Repurpose support: as devices or edge are repurposed, extended, or replace, services change and possible move
- Our Approach
 - Portable, multi-scale, distributed “Functions as a Service” (FaaS)

Cloud Functions as a Service

- Also called “serverless” computing
 - AWS Lambda, Google Functions, Azure Functions
- For cloud
 - Developer writes “stateless” functions in high-level language
 - Packaged with libraries and dependencies
 - Uploaded to FaaS platform in a specific cloud
 - Executed by the cloud in a “container” (no servers required for isolation)
 - Event-driven programming model
 - Functions are only triggered when a specific event occurs (much cheaper!)
 - Storage and communication are through external cloud services



Google Cloud Functions

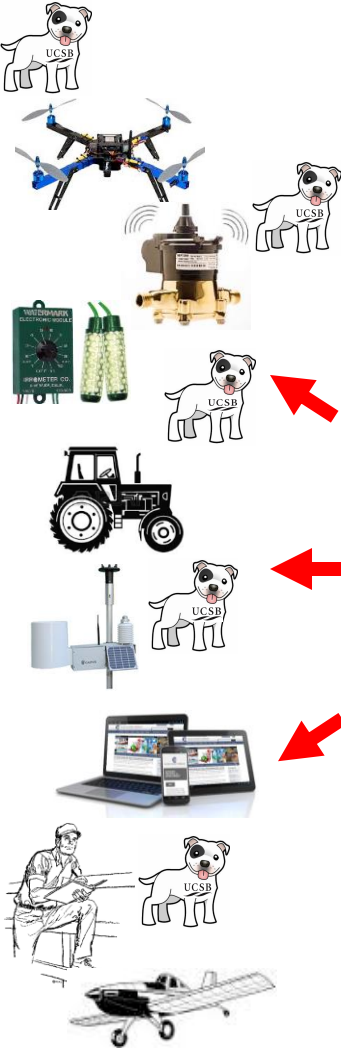
SPOT



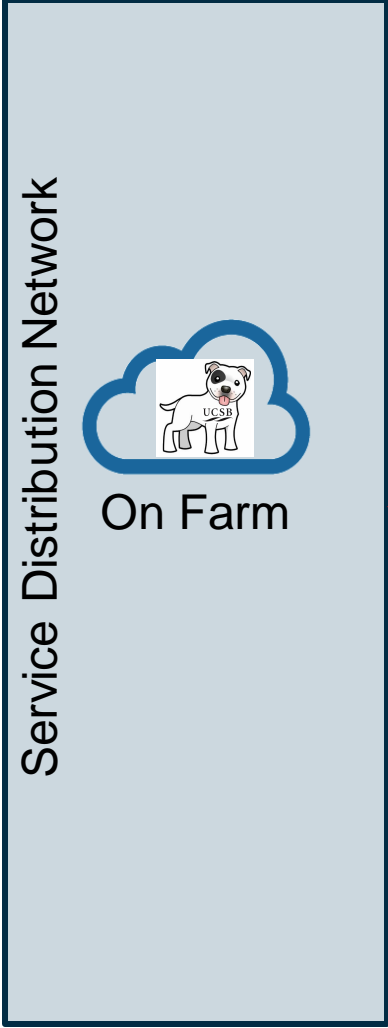
- **S**erverless **P**latform **O**f **T**hings
- Multi-scale, distributed FaaS
 - Run on microcontrollers, small Linux (Raspberry Pi, Intel NUC), Campus Clouds (NSF Aristotle), public clouds (AWS, Azure, Google, IBM)
 - SPOT program ports (source code level) without modification between scales
- SPOT includes
 - Append-only storage abstraction (distributed data durability, eventual consistency)
 - Log-based runtime that tracks causal dependencies (debugging)
- ***SPOT is a way to write multi-scale network-facing services***
- Open Source
 - <https://github.com/MAYHEM-Lab/cspot.git>
 - Join us! Contribute!

UCSB IoT SPOT

Device Tier



Edge Tier



Regional Tier



Public Cloud Tier



Courtesy Chandra Krintz
UCSB Computer Science



Portable and Power Efficient

- Dispatch times for clock read FaaS Handlers
- SPOT is 1 to 2 orders of magnitude faster than AWS Lambda
- Power being proportional to latency

System	Mean (ms)	Stdev (ms)	95% (ms)
Esp8266 ucontroller	38	0.15	40
Raspberry Pi (ARM)	37	6.8	48
Intel NUC	4.0	0.63	4.9
UCSB Cloud	5.0	1.6	7.0
AWS SPOT EC2 in C	5.0	0.96	6.6
AWS SPOT EC2 in Python	18	3.1	23
<i>AWS Lambda</i>	253	90	584

What Will the Brave New World Need?

- One ring to rule them all
 - SPOT is source-code level portable
 - We would love to ship binaries (saves power and time needed to recompile)
 - It would be great if it were “ARM Everywhere”
 - Device, Edge, and Cloud
- Distributed and secure registration and discovery
 - Blockchain?
- High-level platform programming abstractions
 - SPOT is an assembly language for distributed IoT

IoT and Future Cloud

- High-performance Computing (HPC) is really going to matter
 - Predictions about the physical world based on sensor telemetry requires faster-than-real time simulations
- Tiered cloud architecture
 - The data will move only as far from its source as it needs to and no farther
 - Service Distribution Network: like a CDN but for services
- UCSB Institute for Energy Efficiency and ARM (Brent Gorda, HPC) are exploring a research partnership to study these issues
- **Rich's Business Prediction:**
 - *The company that captures this new architecture will be the Microsoft, Intel, or Google of the next decade.*

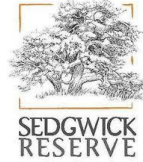


Talent Recruiting: The New Resume'



Thanks!

- Collaborators: UCSB, UCSB IEE, LREC, CalPoly, Fresno State, Powwow Energy, Sedgwick Reserve, Private Growers



- Support: Google, Intel, IBM Research, Microsoft Research, NSF, NIH, California Energy Commission



Chandra Krintz



Rich Wolski

Students:



Fatih Bakir



Kyle Carson



Gareth George



Nevena Golubovic



Carly Larsson



Wei-Tsung Lin



Andy Rosales Elias



Nazmus Saquib



John Thomason



Michael Zhang

UCSB RACELab

The Lab for Research on Adaptive Computing Environments
Computer Science Department, Harold Frank Hall (E-5), Santa Barbara, CA

rich@cs.ucsb.edu, ckrintz@cs.ucsb.edu

