# The Intel Science & Technology Center for Cloud Computing

Phil Gibbons, Co-PI
December 13, 2013

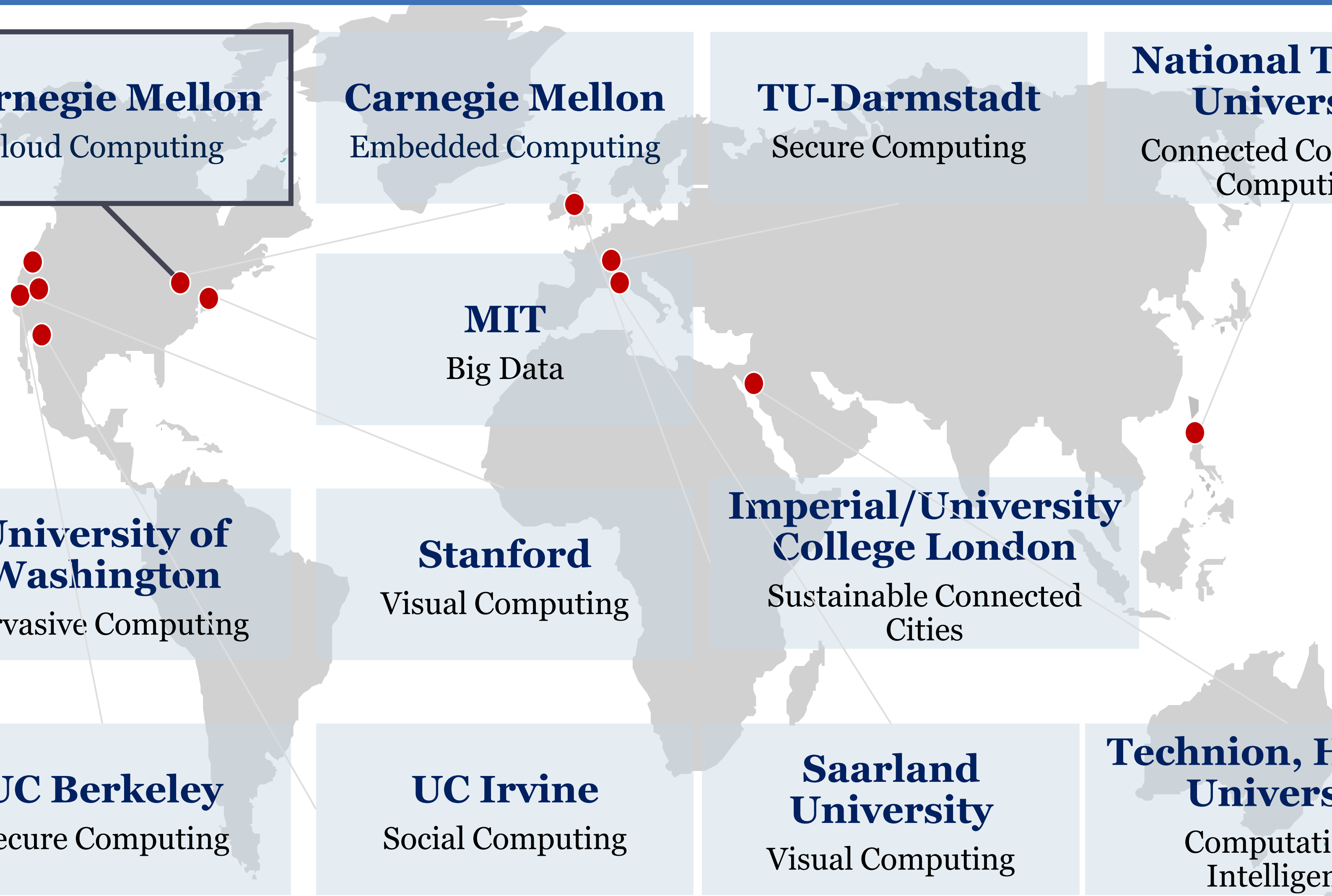
http://www.istc-cc.cmu.edu/

Intel Science & Technology
Center for Cloud Computing

# Abstract (Hidden slide)

The Intel Science and Technology Center (ISTC) for Cloud Computing is a five year, $15M research partnership between Carnegie Mellon, Georgia Tech, Princeton, UC Berkeley, U. Washington, and Intel to research underlying infrastructure enabling the future of cloud computing. Now in its third year, the center has made significant advances in the areas of specialization, automation, big data, and to-the-edge, with 150+ papers, popular open source code releases, and initial tech transfer into Intel. This talk will overview the center's research agenda, highlight some of the key results, and preview where things are headed next. The last part of the talk will provide a deeper dive into the center's research on machine learning over big data ("Big Learning").

# Intel Science & Technology Centers (ISTC)
# Intel Collaborative Research Institutes (ICRI)



**Carnegie Mellon**
Cloud Computing

**Carnegie Mellon**
Embedded Computing

**TU-Darmstadt**
Secure Computing

**National Taiwan University**
Connected Contextual Computing

**MIT**
Big Data

**University of Washington**
Pervasive Computing

**Stanford**
Visual Computing

**Imperial/University College London**
Sustainable Connected Cities

**UC Berkeley**
Secure Computing

**UC Irvine**
Social Computing

**Saarland University**
Visual Computing

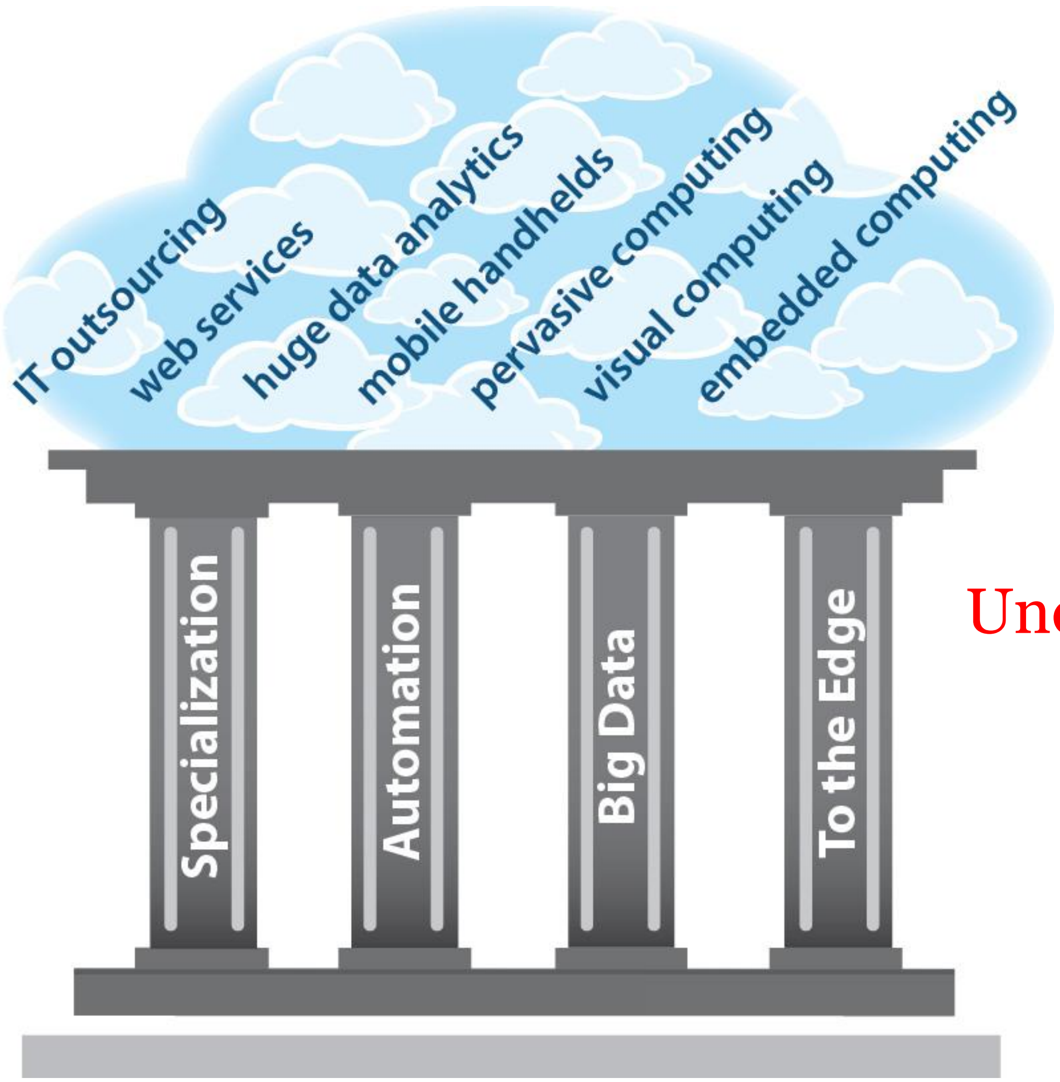
**Technion, Hebrew University**
Computational Intelligence

Open IP, Open Pubs, Open Source. Typically, 3+2 years

# ISTC for Cloud Computing

$11.5M over 5 years + 4 Intel researchers. Launched Sept 2011

25 faculty
87 students




Underlying Infrastructure enabling the future of cloud computing

www.istc-cc.cmu.edu

# ISTC for Cloud Computing: Faculty

- Carnegie Mellon University
  - Greg Ganger (PI), Dave Andersen, Guy Blelloch, Garth Gibson, Mor Harchol-Balter, Todd Mowry, Onur Mutlu, Priya Narasimhan, M. Satyanarayanan, Dan Siewiorek, Alex Smola, Eric Xing
- Georgia Tech
  - Greg Eisenhower, Ada Gavrilovska, Ling Liu, Calton Pu, Karsten Schwan, Matthew Wolf, Sudha Yalamanchili
- Princeton University
  - Mike Freedman, Margaret Martonosi
- University of California at Berkeley
  - Anthony Joseph, Randy Katz, Ion Stoica
- University of Washington
  - Carlos Guestrin
- Intel Labs
  - Phil Gibbons (PI), Michael Kaminsky, Mike Kozuch, Babu Pillai

Carnegie Mellon University

Georgia Tech

PRINCETON UNIVERSITY

UC Berkeley

UNIVERSITY of WASHINGTON

intel

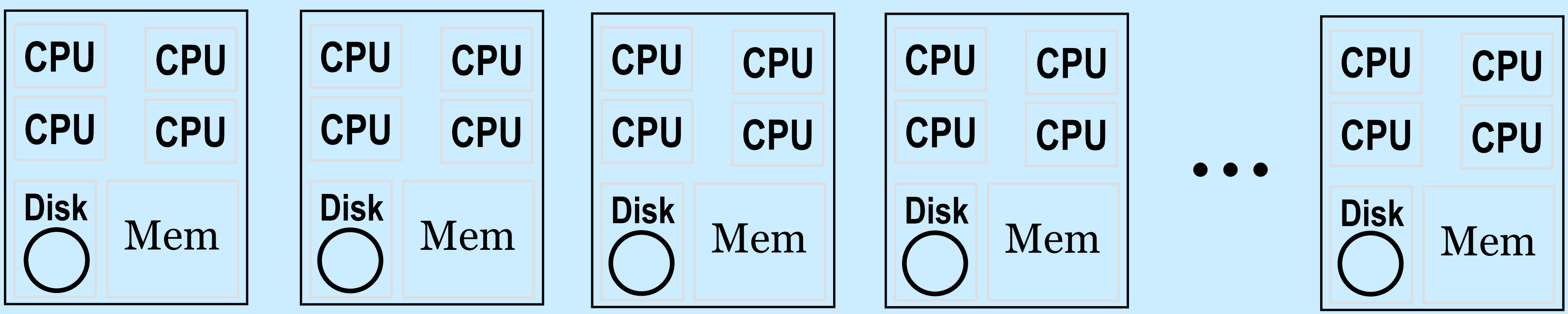
# Outline

- Highlights from 4 Research Pillars
  - Specialization
  - Automation
  - Big Data
  - To the Edge

- Deeper dive on Big Learning

# Cloud Computing & Homogeneity

- Traditional data center goal: Homogeneity
  + Reduce administration costs: maintenance, diagnosis, repair
  + Ease of load balancing



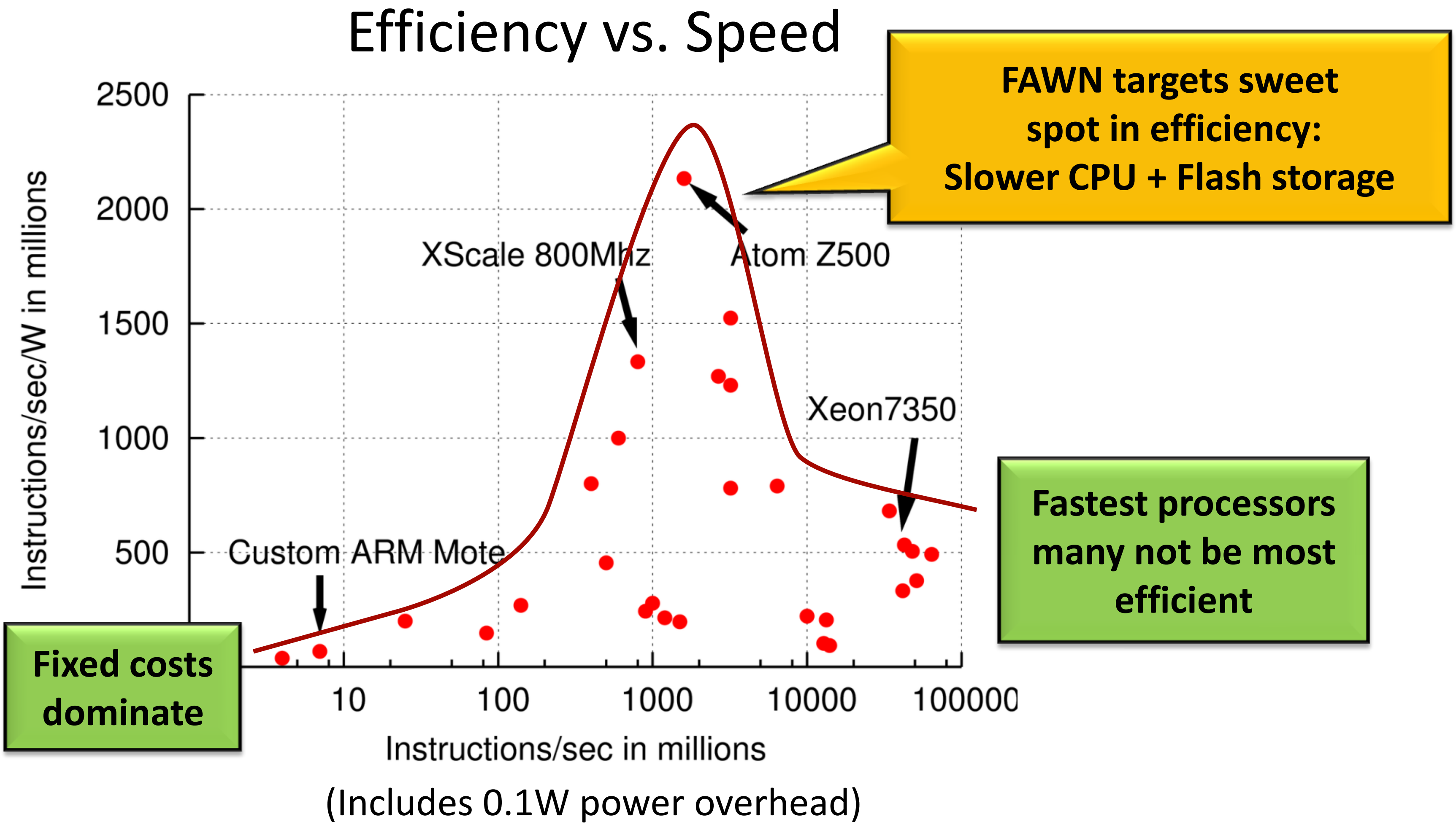

Ideal: single Server Architecture tailored to the workload

| Specialization | Automation | Big Data | To the Edge |
|---|---|---|---|

# Homogeneity: Challenges

- No single workload: Mix of customer workloads
  - Computation-heavy apps (powerful CPUs, little I/O BW)
  - Random I/O apps (I/O latency bound)
  - Streaming apps (I/O BW bound, little memory)
  - Memory-bound apps
  - Apps exploiting hardware assists such as GPUs
- Common denominator Server Architecture falls short
  - E.g., Two orders of magnitude loss in energy efficiency

# Targeting the Sweet Spot in Energy Efficiency



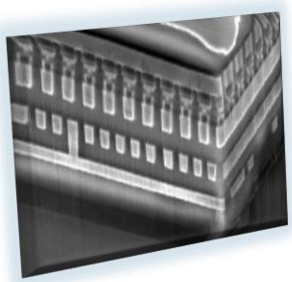

* Numbers from spec sheets

[FAWN: A Fast Array of Wimpy Nodes, Andersen et al, SOSP'09]

# Specialization Pillar

*Low power nodes*

*Many-core*



*Phase-change memory (PCM)*

- Specialization is fundamental to efficiency
  - No single platform best for all application types
  - Called **division of labor** in sociology
- Cloud computing must embrace specialization
  - As well as consequent heterogeneity and change-over-time
  - Stark contrast to common cloud thinking
- New approaches needed to enable…
  - Effective mixes of targeted and general platform types, heterogeneous multi-cores, hybrid memories

| Specialization | Automation | Big Data | To the Edge |
|---|---|---|---|

# Specialization Projects

- **S1: Specialized Platforms of Wimpy Nodes**
  - exploring + extending range of apps that run (most) efficiently on such platforms by overcoming OS limits, memory limits, and scalability issues

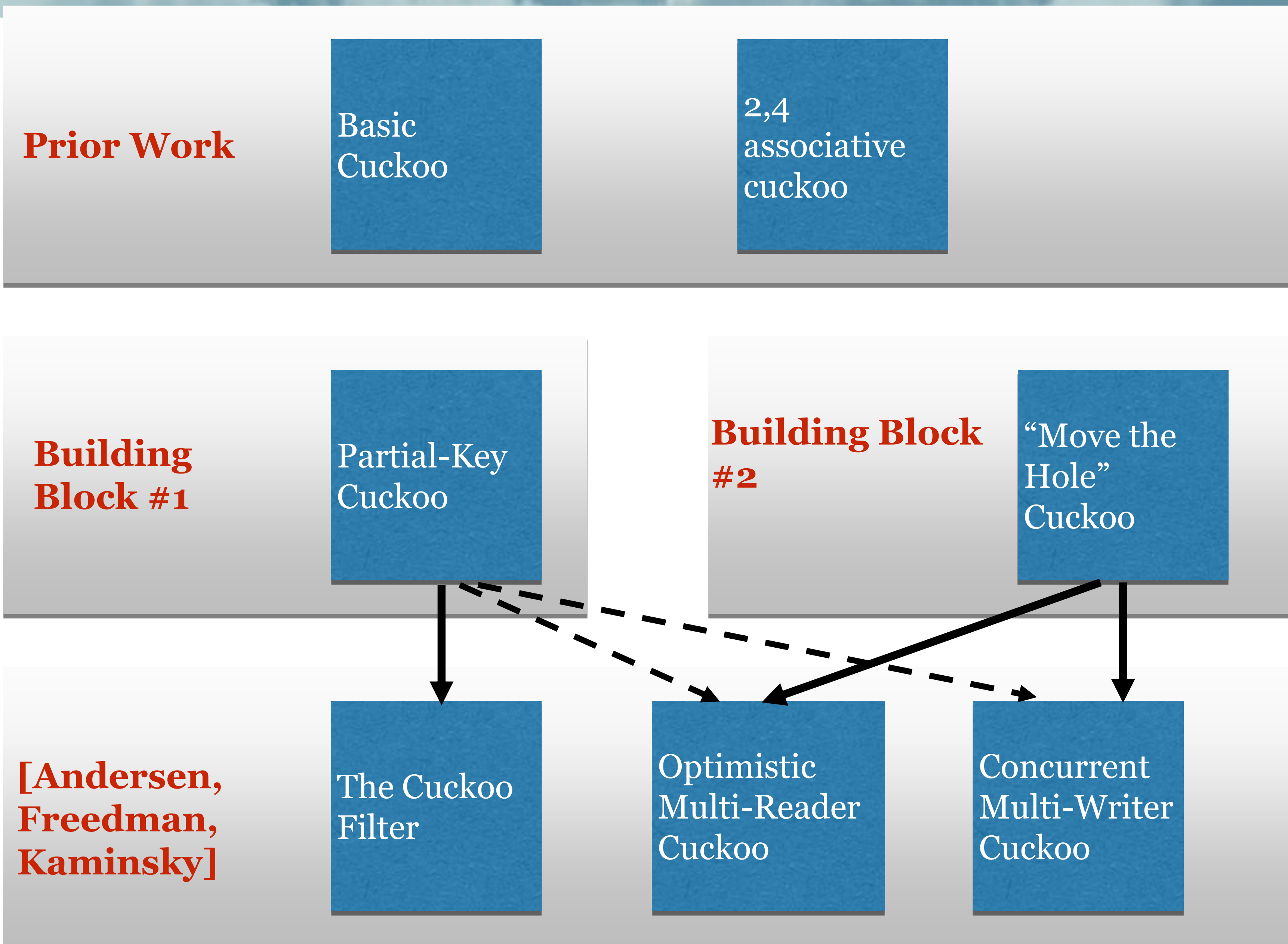
- **S2: Specialized Platforms of Heterogeneous Multi-Cores**
  - exploring best ways to devise and use heterogeneity on multi-core nodes, considering core types, accelerators, DRAM/NVM memory, frequency scaling, and sleep states, with a focus on cloud's virtualized, multi-tenancy workloads

| Specialization | Automation | Big Data | To the Edge |
|---|---|---|---|

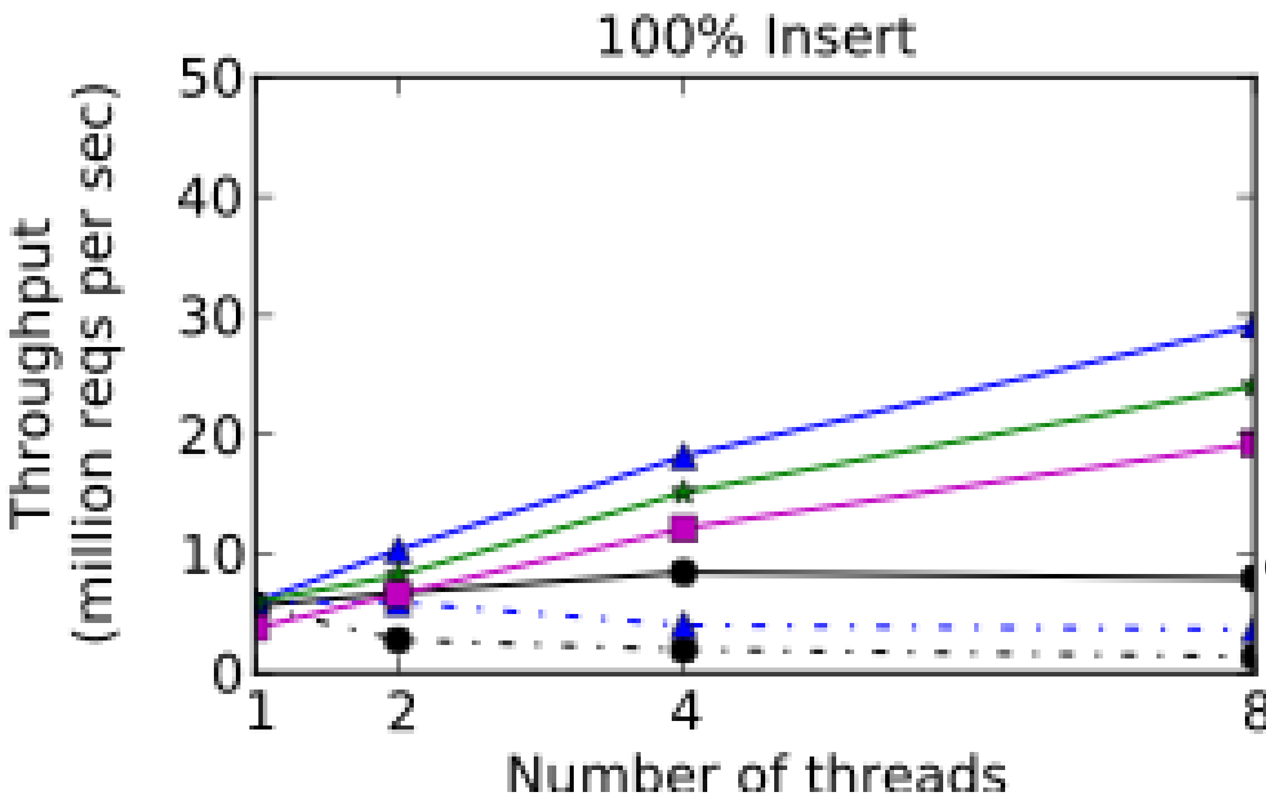
# Specialization Highlights

- Selected Research Highlights
  - **SILT: A Memory-Efficient, High-Performance Key-Value Store**, Andersen, Kaminsky, SOSP'11
    - key-value store design with very memory-efficient, scalable indices, combined with model-driven tuning to match workload
  - **Staged Memory Scheduling: Achieving High Performance and Scalability in Heterogeneous Systems**, Multu, ISCA'12
    - new memory controller design that enhances performance, reduces interference, and increases fairness for apps running on distinct heterogeneous cores (e.g., GPUs and CPUs)
  - **The Forgotten 'Uncore': On the Energy-Efficiency of Heterogeneous Cores**, Schwan, Usenix ATC'12
    - investigates the opportunities and limitations in using heterogeneous multicore processors to gain energy-efficiency, highlighting the importance of the "uncore" subsystem shares by all cores to such goals

# Fast, Memory Efficient (Cuckoo) Hashing



4 core
Haswell
Desktop

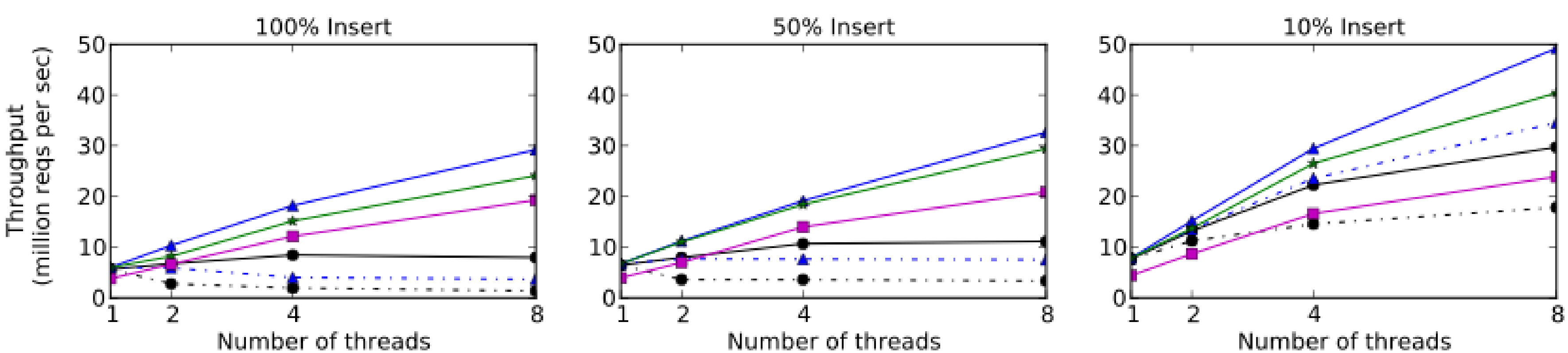



Cuckoo-TSX

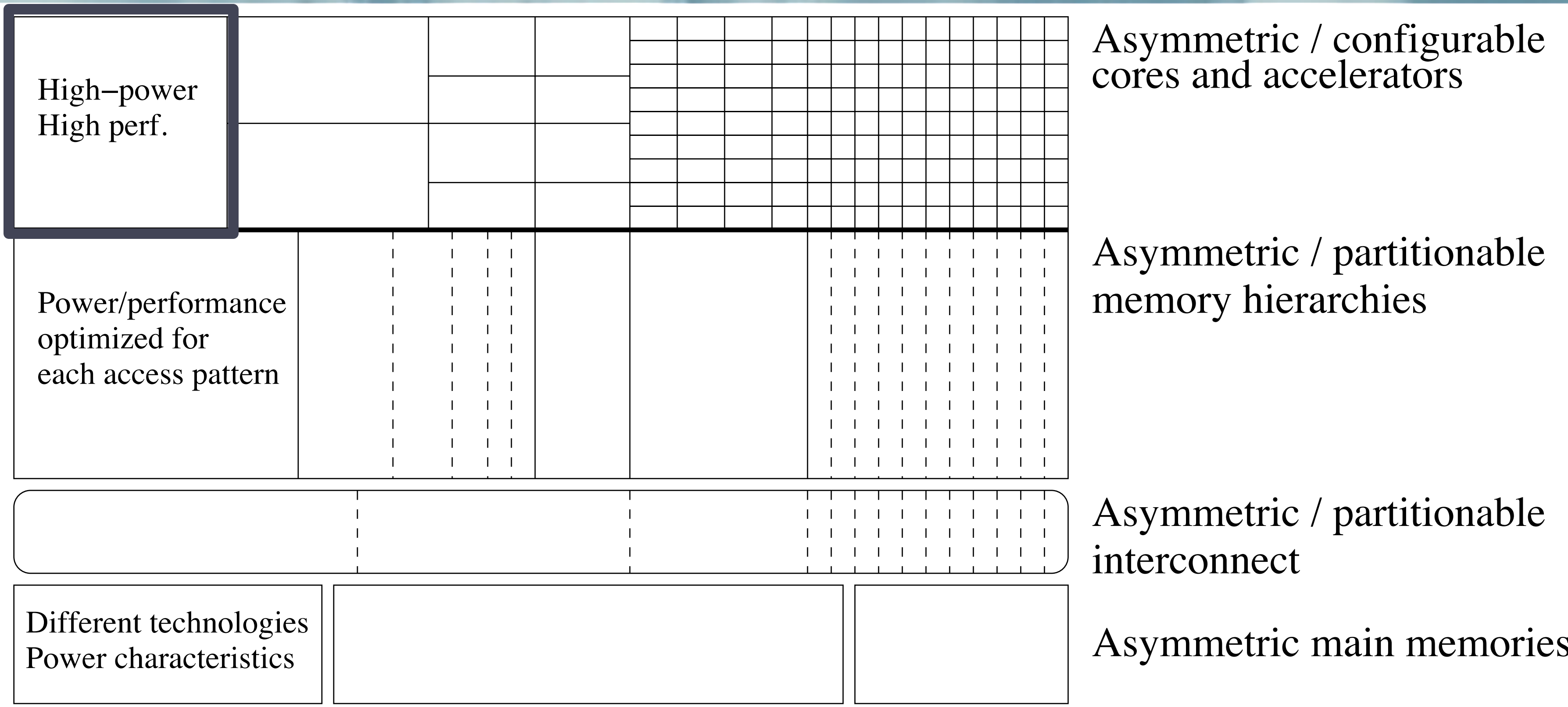
Cuckoo-Spinlock

TBB hash_map

Cuckoo-opt-global

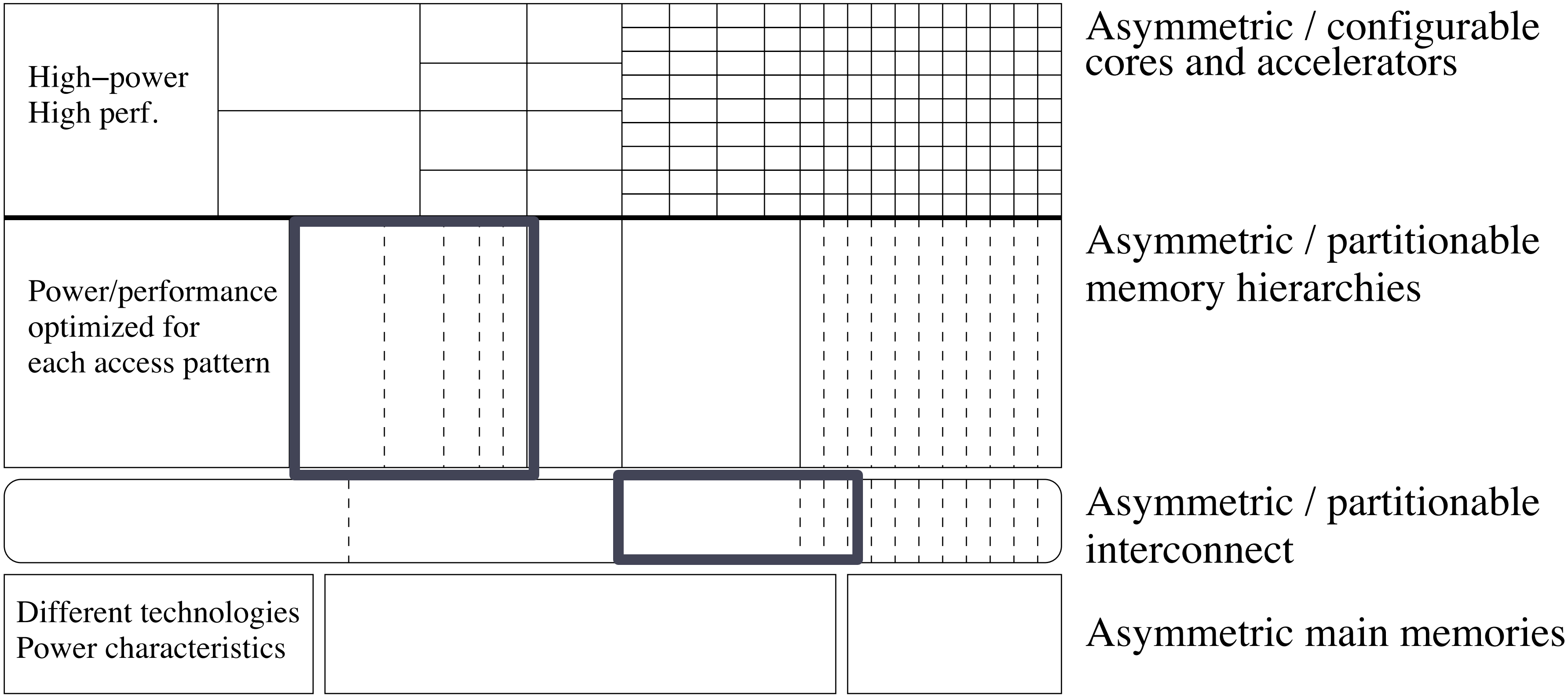


# Exploiting Heterogeneity (1)



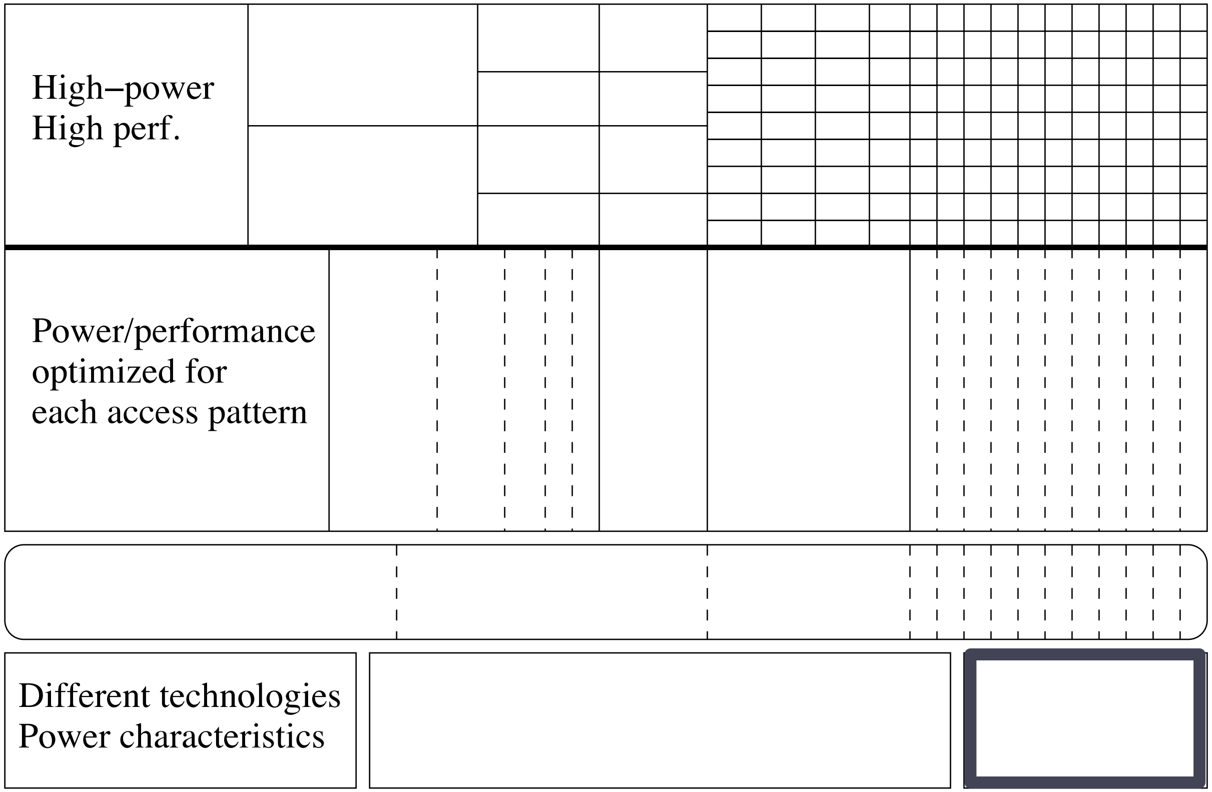

- Execute critical/serial sections on high-power, high-performance cores/resources [Suleman+ ASPLOS'09, ISCA'10, Top Picks'10'11, Joao+ ASPLOS'12]
  - Programmer can write less optimized, but more likely correct programs

# Exploiting Heterogeneity (2)




- Partition memory controller and on-chip network bandwidth asymmetrically among threads [Kim+ HPCA 2010, MICRO 2010, Top Picks 2011] [Nychis+ HotNets 2010] [Das+ MICRO 2009, ISCA 2010, Top Picks 2011]
  - Higher performance and energy-efficiency than symmetric/free-for-all

# Exploiting Heterogeneity (3)




Asymmetric / configurable cores and accelerators

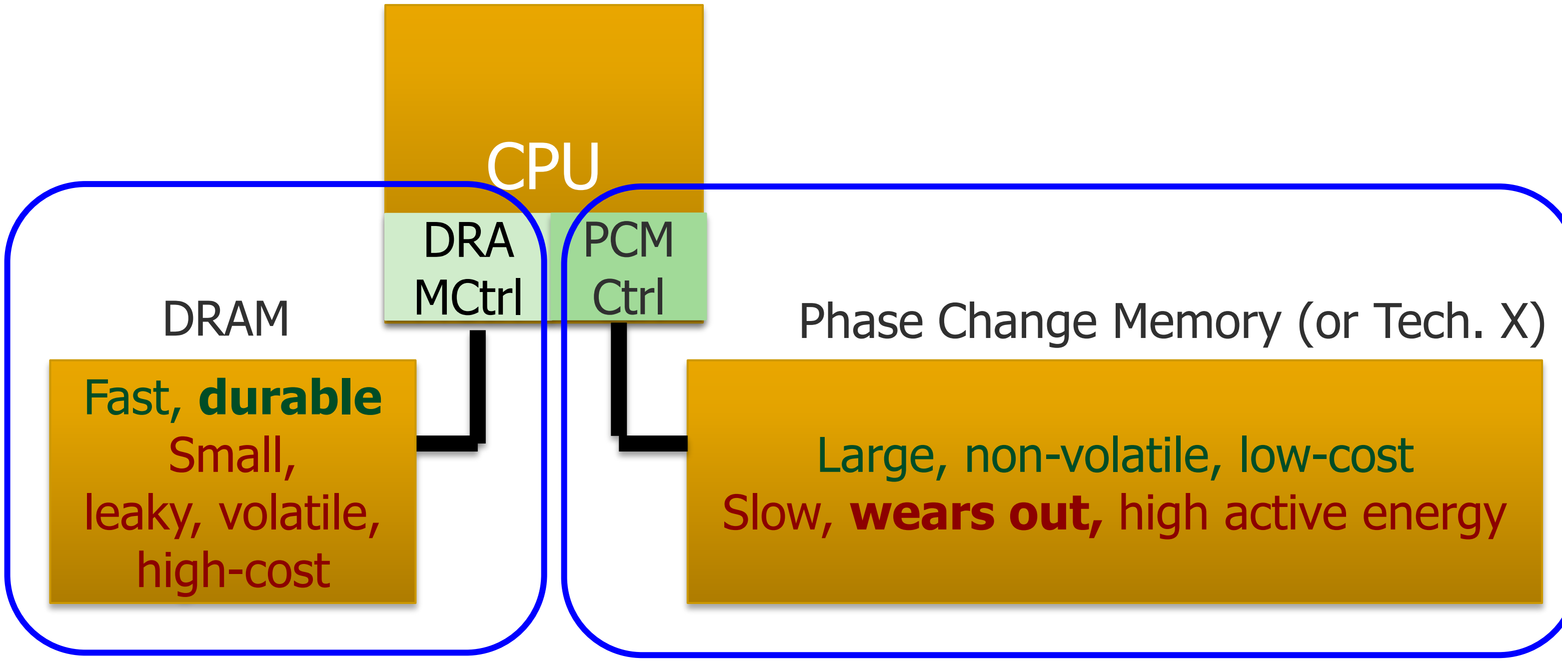
Asymmetric / partitionable memory hierarchies

Asymmetric / partitionable interconnect

Asymmetric main memories

- Have multiple different memory scheduling policies; apply them to different sets of threads based on thread behavior [Kim+ MICRO 2010, Top Picks 2011] [Ausavarungnirun, ISCA 2012]
  - Higher performance and fairness than a homogeneous policy

# Hybrid Memory Systems




Hardware/software manage data allocation and movement
to achieve the best of multiple technologies

Meza+, "Enabling Efficient and Scalable Hybrid Memories," IEEE Comp. Arch. Letters, 2012.

Yoon, Meza et al., "Row Buffer Locality Aware Caching Policies for Hybrid Memories," ICCD 2012 Best Paper Award.

# Automation Pillar

- Automation is crucial to cloud reaching potential
  - We suspect that no one here needs to be convinced of this…
- Management is very hard, but cloud makes it worse
  - Much larger scale
  - Much more varied mix of applications/activities
  - Much less pre-knowledge of applications
  - And, we're adding in platform specialization ☺
- Leaps forward needed on many fronts…
  - Diagnosis, scheduling, instrumentation, isolation, tuning, …

| Specialization | **Automation** | Big Data | To the Edge |
| --- | --- | --- | --- |

# Automation Projects

- **A1: Resource Scheduling for Heterogeneous Cloud Infrastructures**
  - maximizing the effectiveness of a cloud composed of diverse specialized platforms servicing diverse app types
  - enabling software framework specialization via hierarchical scheduling
- **A2: Problem Diagnosis and Mitigation**
  - new tools and techniques for rapid, robust diagnosis of failures and performance problems
  - automated mitigation based on "quick and dirty" online diagnoses

# Automation Highlights

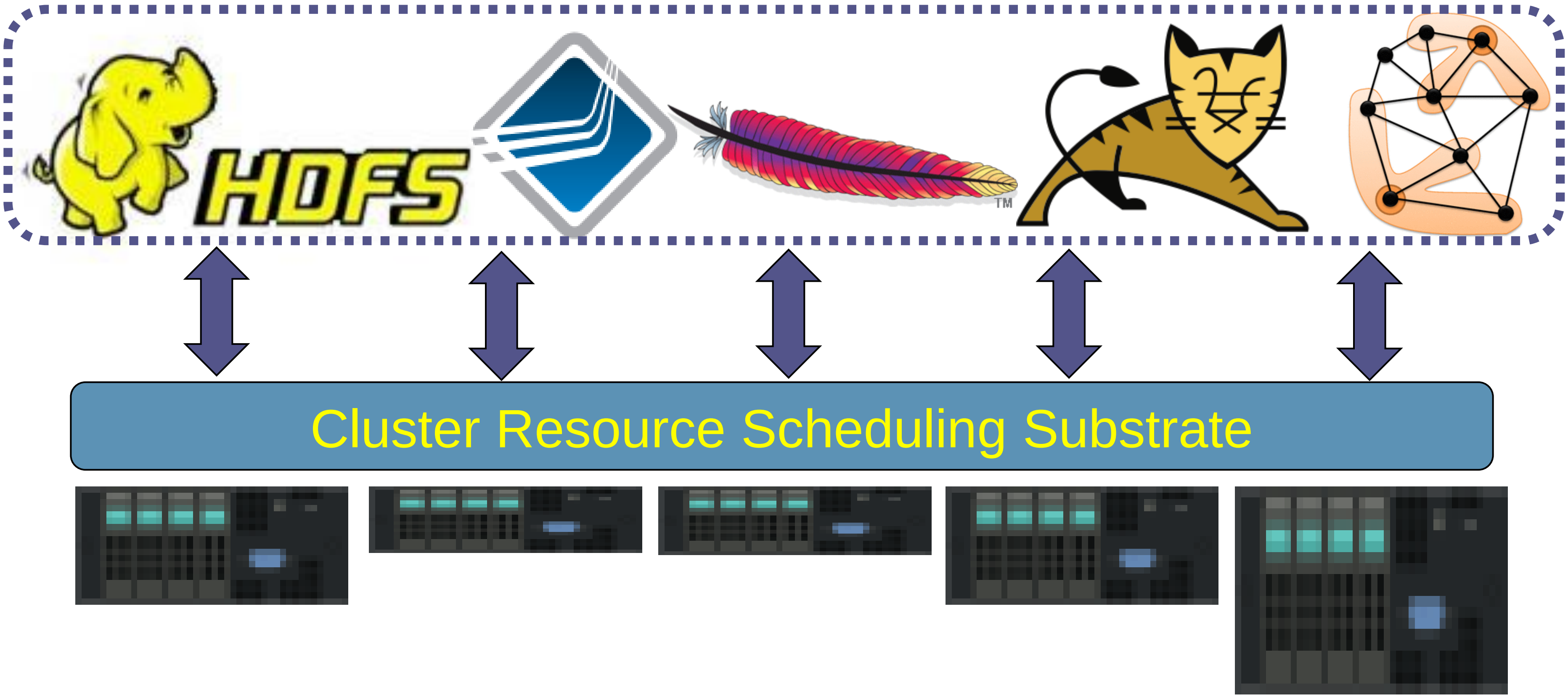
- Selected Research Highlights
  - **Energy Efficiency for Large-Scale MapReduce Workloads with Significant Interactive Analysis**, Katz, EuroSys'12
    - Energy efficient MapReduce workload manager motivated by empirical analysis of real-life MapReduce Interactive Analysis traces
  - **Are Sleep States Effective in Data Centers?**, Harchol-Balter, Kozuch, IGCC'12
    - Quantifies the benefits of sleep states across three dimensions: (i) the variability in the workload trace, (ii) the type of dynamic power management policy employed, and (iii) the size of the data center
  - **Reliable State Monitoring in Cloud Datacenters**, Liu, CLOUD'12
    - Quantitatively estimates the accuracy of monitoring results to capture uncertainties introduced by messaging dynamics, and adapts to non-transient messaging issues by reconfiguring monitoring algorithms

# Automation Highlights

- Selected Research Highlights
  - **Hierarchical Scheduling for Diverse Datacenter Workloads**, Stoica, SOCC'13
    - Dominant Resource Fairness (NSDI'11) extended to hierarchical setting
  - **Sparrow: Distributed, Low Latency Scheduling**, Stoica, SOSP'13
    - Decentralized scheduler for jobs with low-latency (100 ms) parallel tasks
  - **A Hidden Cost of Virtualization when Scaling Multicore Applications**, G., Kozuch, HotCloud'13
    - Idleness consolidation to reduce a surprising VMM cost
  - **Guardrail: A High Fidelity Approach to Protecting Hardware Devices from Buggy Drivers**, G., Kozuch, Mowry, ASPLOS'14

# Scheduling for Heterogeneous Clouds

- Many execution frameworks + Mix of platform types
- Goal: Cluster Scheduler that gets frameworks to “play nice” & matches work to suitable platform

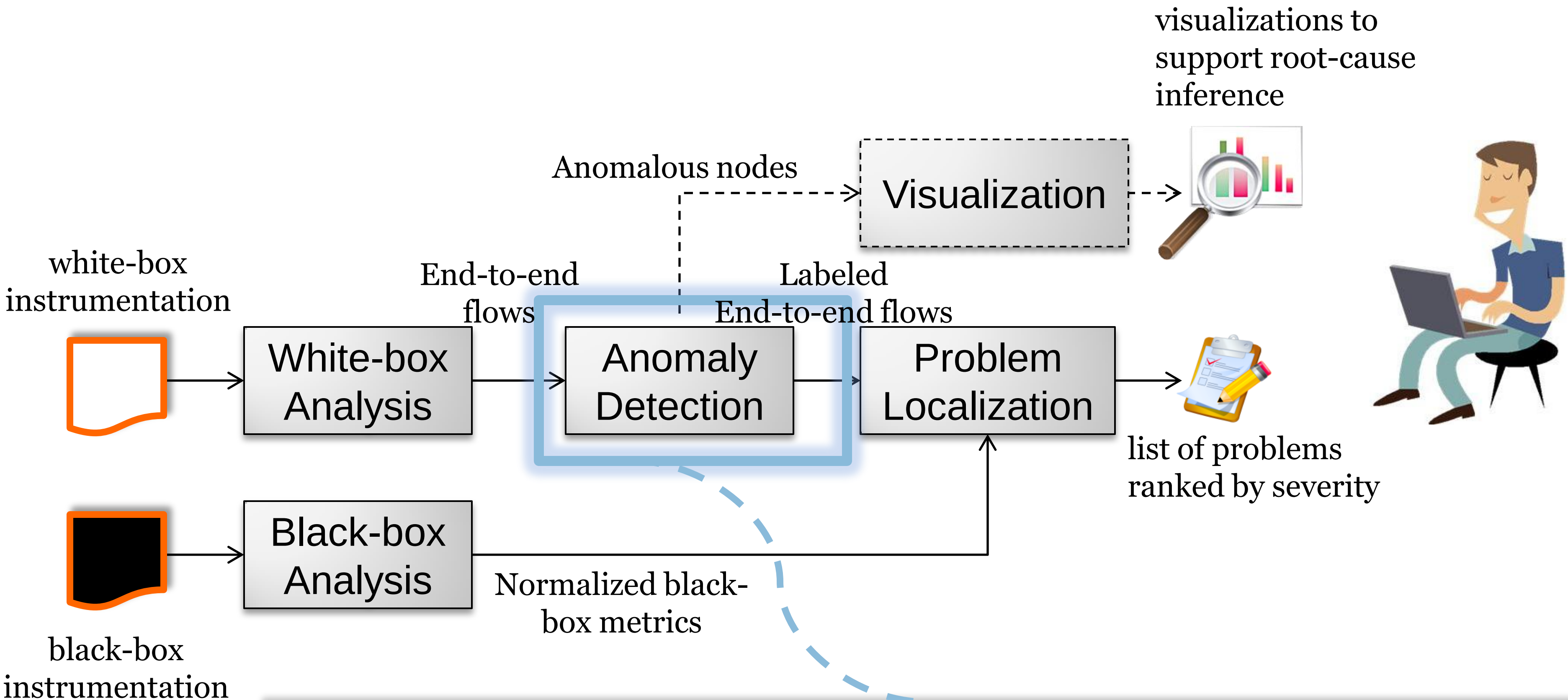


# Scheduling for Heterogeneous Clouds

- **Mesos: A platform for fine-grained resource sharing in the data center**, Joseph, Katz, Stoica, NSDI'11

- **Tetrisched: Space-Time Scheduling for Heterogeneous Datacenters**, Ganger, Kozuch, Harchol-Balter
  - Extends Mesos' resource offer to utility function; tetris-inspired scheduler

# Anomaly Detection in Hadoop Clusters



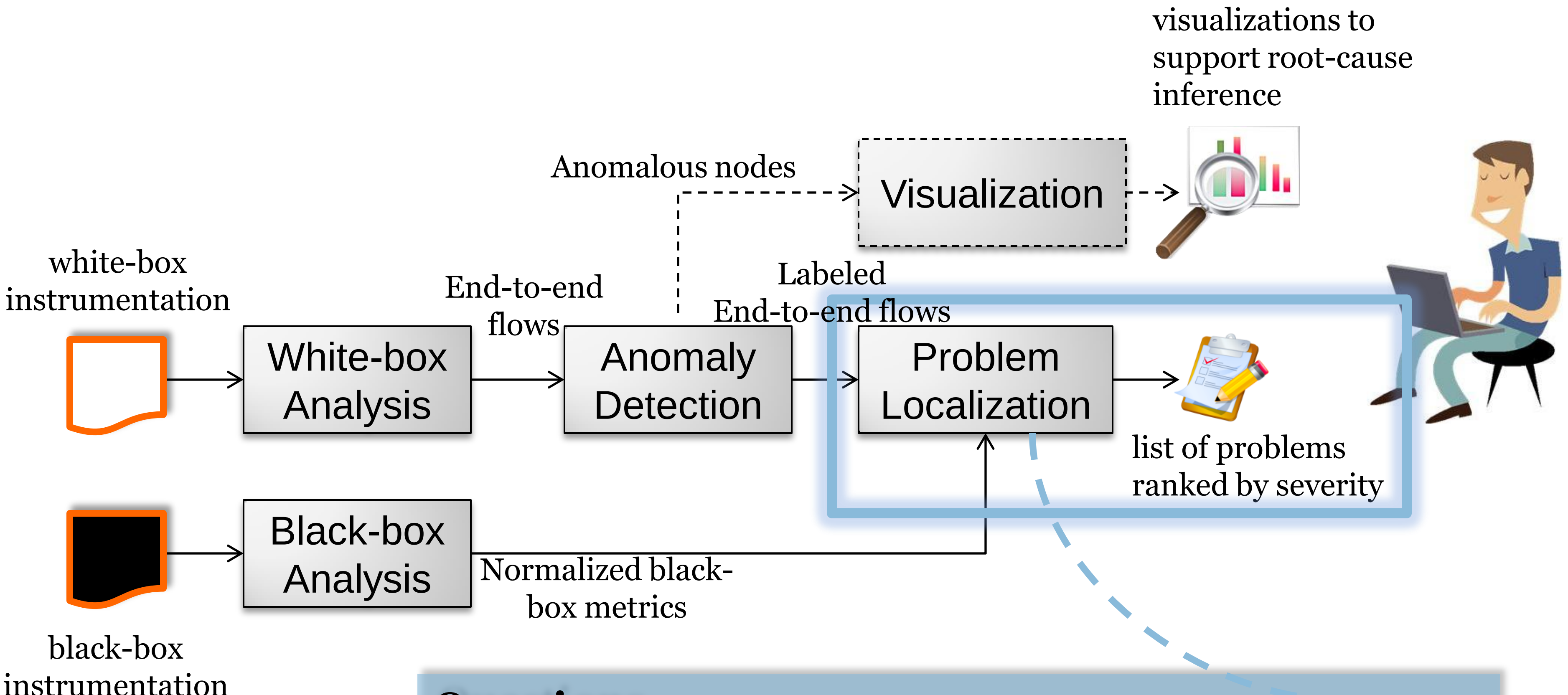

**Questions**

- How to detect performance problems in the absence of labeled data?
- How to distinguish legitimate application behavior vs. problems?

# Anomaly Detection -- Approach

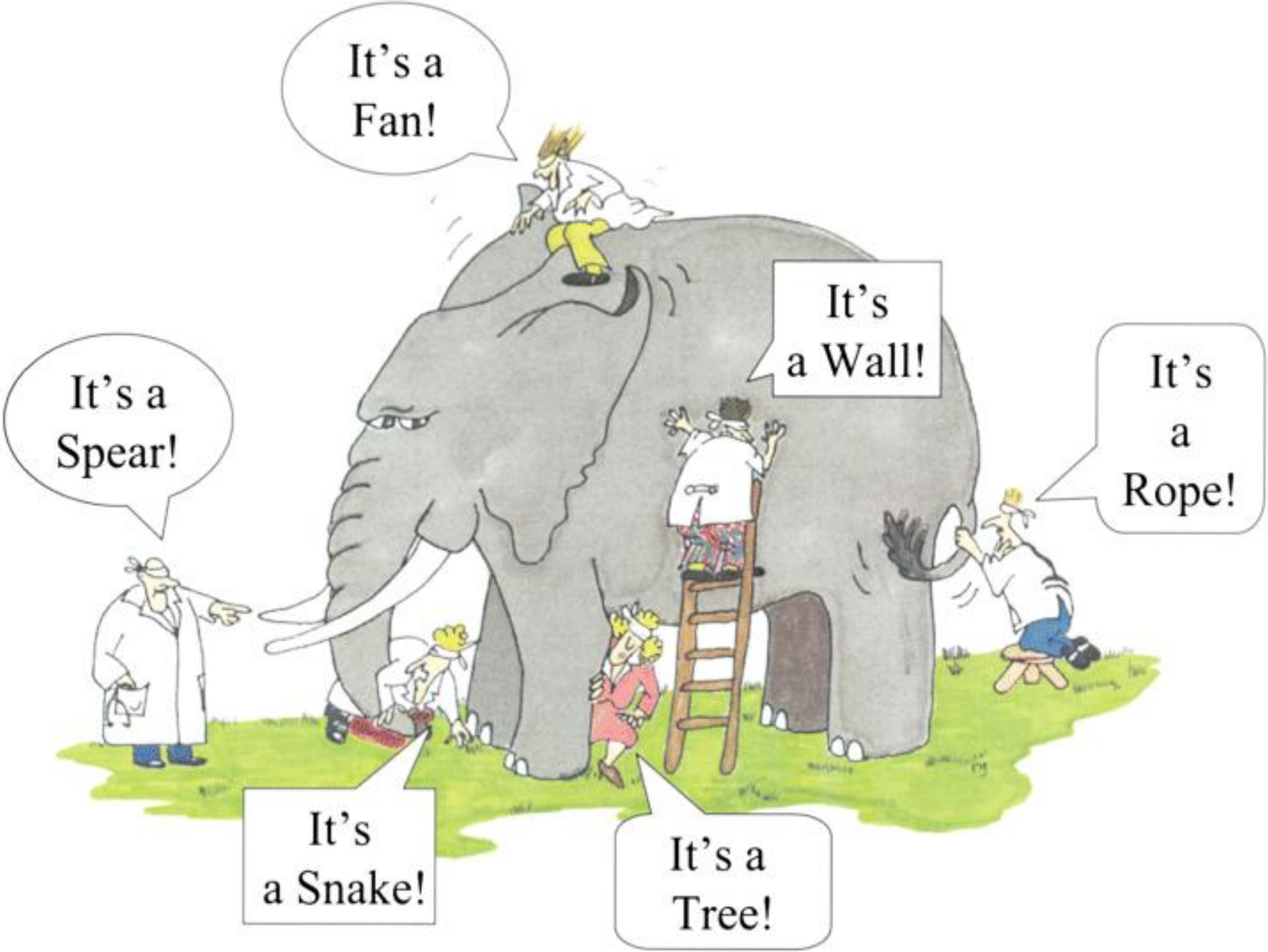
- **Detect performance problems using "peers"**
  - Empirical analysis of production data to identify peers
    - 219,961 successful jobs (Yahoo! M45 and OpenCloud)
    - 89% of jobs had low variance in their Map durations
    - 65% of jobs had low variance in their Reduce durations
  - Designate tasks belonging to the same job as peers
- At the same time, behavior amongst peers can legitimately diverge due to various application factors
  - Identified 12 such factors on OpenCloud
  - Example: HDFS bytes written/read

# Problem Localization




**Questions**

- How to identify problems due to combination of factors?
- How to distinguish multiple ongoing problems?
- How to find resource that caused the problem?
- How to handle “noise” due to flawed anomaly detection?


It's a Fan!
It's a Wall!
It's a Spear!
It's a Rope!
It's a Snake!
It's a Tree!

# Fusing the Metrics

# Impact of Fusion

**QUESTION:** Does fusion of metrics provide insight on root-cause?

**METHOD:** Hadoop EC2 cluster, 10 nodes, fault injection.

- Apply problem localization with fused white/black-box metrics.

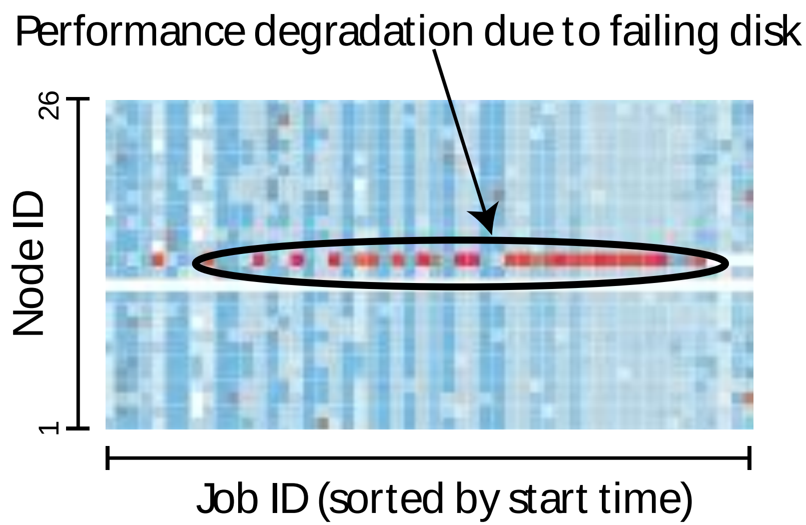
| | Top Metrics Indicted | | Insight on root-cause |
|---|---|---|---|
| **Fault Injected** | **White box** | **Black-box** | |
| Disk hog | Maps | Disk | ✓ |
| Packet-loss | Shuffles | - | ✗ |
| Map hang (Hang1036) | Maps | - | ✓ |
| Reduce hang (Hang1152) | Reduces | - | ✓ |

Fusion of metrics provides insight on most injected faults

# Theia: Visual Signatures of Problems

- Maps anomalies observed to broad problem classes
  - Hardware failures, application issue, data skew
- Supports interactive data exploration
  - Users drill-down from cluster- to job-level displays
  - Hovering over the visualization gives more context
- Compact representation for scalability
  - Can support clusters with 100s of nodes

*USENIX LISA 2012 Best Student-Paper Award



# Big Data Pillar

*Customer Database*

~600 TB

*HD Internet Video*

12 EB/yr

*Particle Physics*

300 EB/yr

- Extracting insights from large datasets
  - "Analytics" or "Data-intensive computing"
  - Becoming critical in nearly every domain
    - likely to dominate future cloud data centers
- Need right programming/execution models
  - For productivity, efficiency, and agility
  - Resource efficient operation on shared, specialized infrastructures


*Estimating the Exaflood,* Discovery Institute, January 2008
*Amassing Digital Fortunes, a Digital Storage Study*, Consumer Electronic Association, March 2008

# Big Data Projects

- **B1: Big Learning Systems**
  - new programming abstractions and execution frameworks enabling efficiency and productivity for large-scale Machine Learning
- **B2: Big Data Storage**
  - exploring trade-offs and new approaches in Big Data storage, including support for high ingress and multi-framework sharing of data

# Big Data Highlights

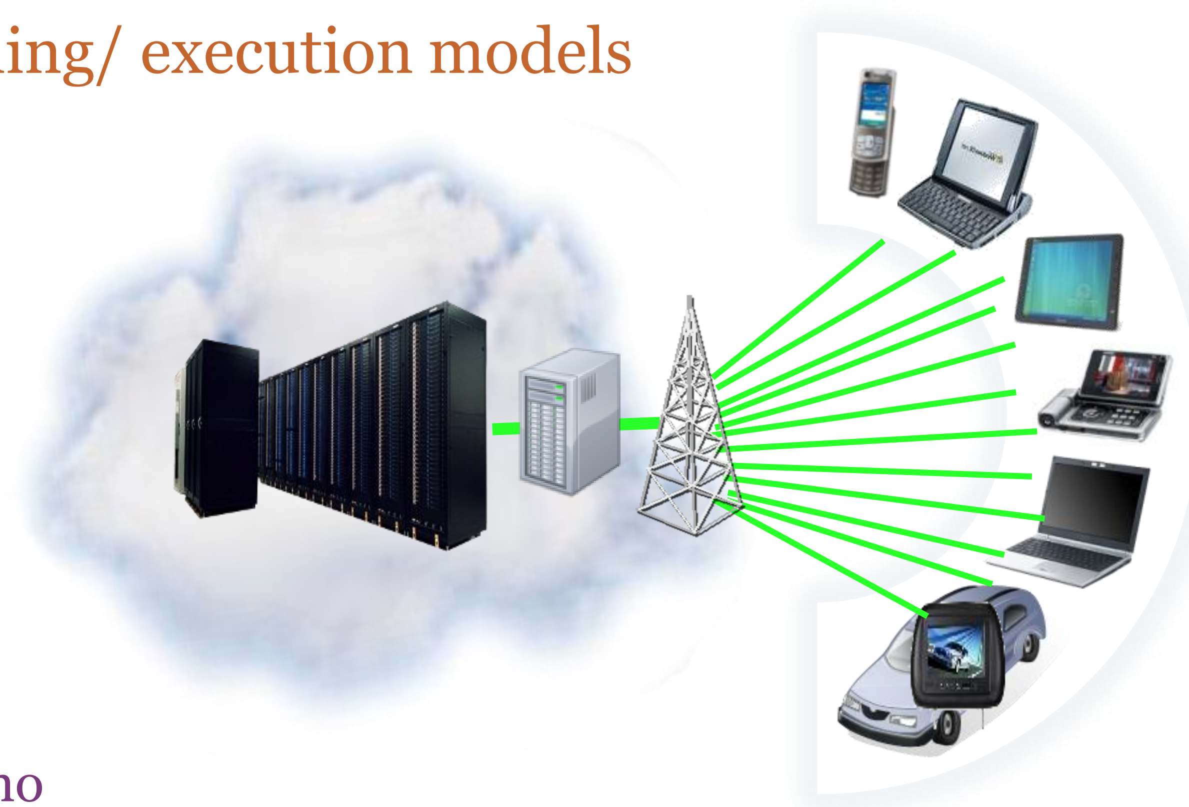
- Selected Research Highlights
  - **LazyBase: Trading Freshness for Performance in a Scalable Database,** Ganger, EuroSys'12
    - Simultaneously ingest atomic batches of updates at a very high throughput and offer quick read queries to a stale-but-consistent version of the data
  - **YCSB++: Benchmarking and Performance Debugging Advanced Features in Scalable Table Stores**, Gibson, SOCC'11
    - Understanding and debugging the performance of advanced features such as ingest speed-up techniques and function shipping filters
  - **Parrot: A Practical Runtime for Deterministic, Stable, and Reliable Threads**, Gibson, SOSP'13

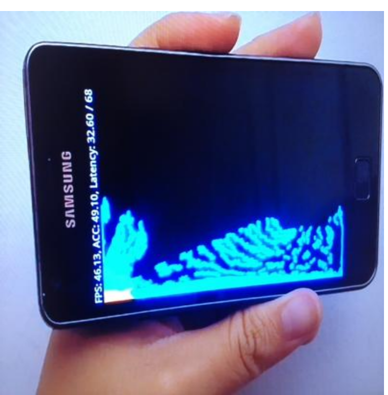
+ Big Learning highlights covered in deeper dive

# To the Edge Pillar

- Edge devices will participate in cloud activities
  - Serving as bridge to physical world (sense/actuate)
  - Enhancing interactivity despite location / connectivity

- Need new programming/ execution models
  - For adaptive cloud
    + edge cooperation






Cloudlet demo

# To the Edge Projects

- **E1: Cloud-Assisted Mobile Client Computations**
  - new abstractions and system architectures for dynamic exploitation of edge-local cloud resources to enable rich edge device experiences

- **E2: Geographically Distributed Data Storage**
  - new techniques for geographically distributed data storage/caching that reduce both access latency & reliance on expensive WAN-uplink bandwidth, while providing the desired scalability, fault tolerance, consistency & findability

| Specialization | Automation | Big Data | **To the Edge** |
|---|---|---|---|

# To the Edge Highlights

- Selected Research Highlights
  - **Don't Settle for Eventual: Stronger Consistency for Wide-Area Storage with COPS**, Andersen, Freedman, Kaminsky, SOSP'11
    - Define Causal+ consistency, with scalable implementation
  - **Stronger Semantics for Low-Latency Geo-Replicated Storage,** Andersen, Freedman, Kaminsky, NSDI'13
    - Eiger improves COPS for read-only, write-only transactions
  - **There Is More Consensus In Egalitarian Parliaments,** Andersen, Freedman, Kaminsky, SOSP'13
    - ePaxos demonstrates significant latency improvement over well-studied Paxos for wide-area replica consistency

# To the Edge Highlights

- Selected Research Highlights
  - **The Impact of Mobile Multimedia Applications on Data Center Consolidation,** Satya, IC2E'13
    - Quantitative support for Cloudlets for multimedia apps
  - **Scalable Crowd-Sourcing of Video from Mobile Devices,** Satya, Mobisys'13
    - Cloudlets store videos locally, send only metadata to backend search engine
  - **Just-in-Time Provisioning for Cyber Foraging,** Satya, Mobisys'13
    - Launch Personalized VM in Cloudlet in 10 seconds, not 5 minutes

# Cloudlets: Bring the cloud to the user

- Provide cloud-like resources, compute services with *logical proximity* to user

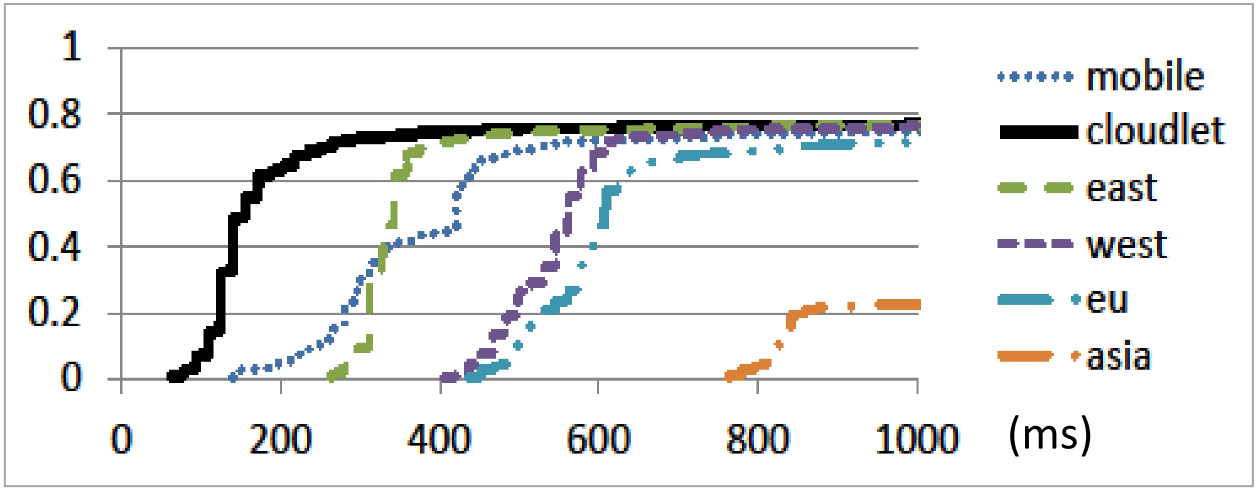
- Like web caches – deployed at the edges

- Like WiFi – decentralized, minimally managed deployments

# Cloudlets vs. On client vs. Cloud

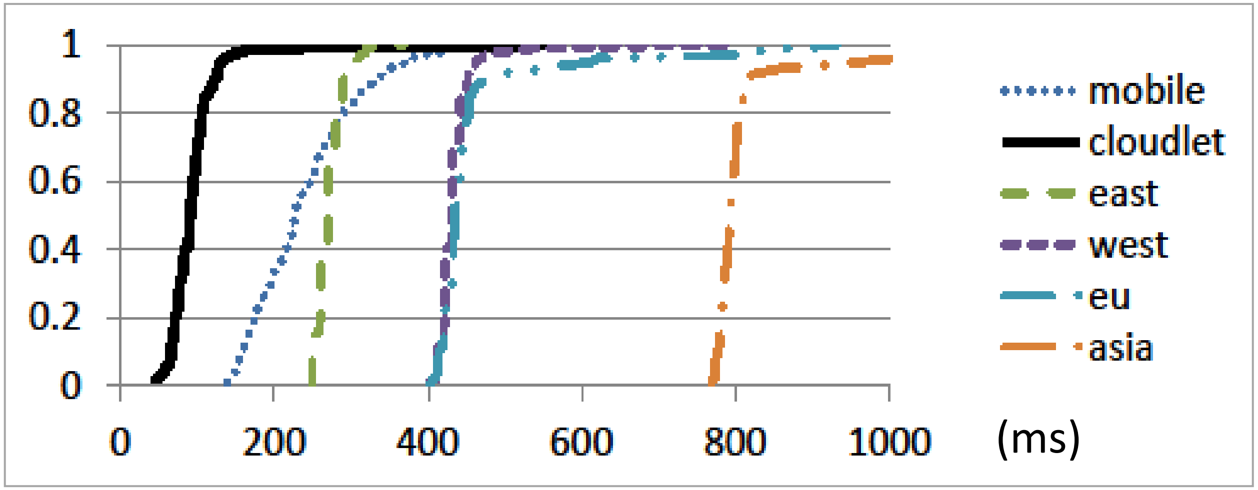
**Face Recognition**

- CDF of 300 requests (images)




**Augmented Reality**

- CDF of 100 requests (images)

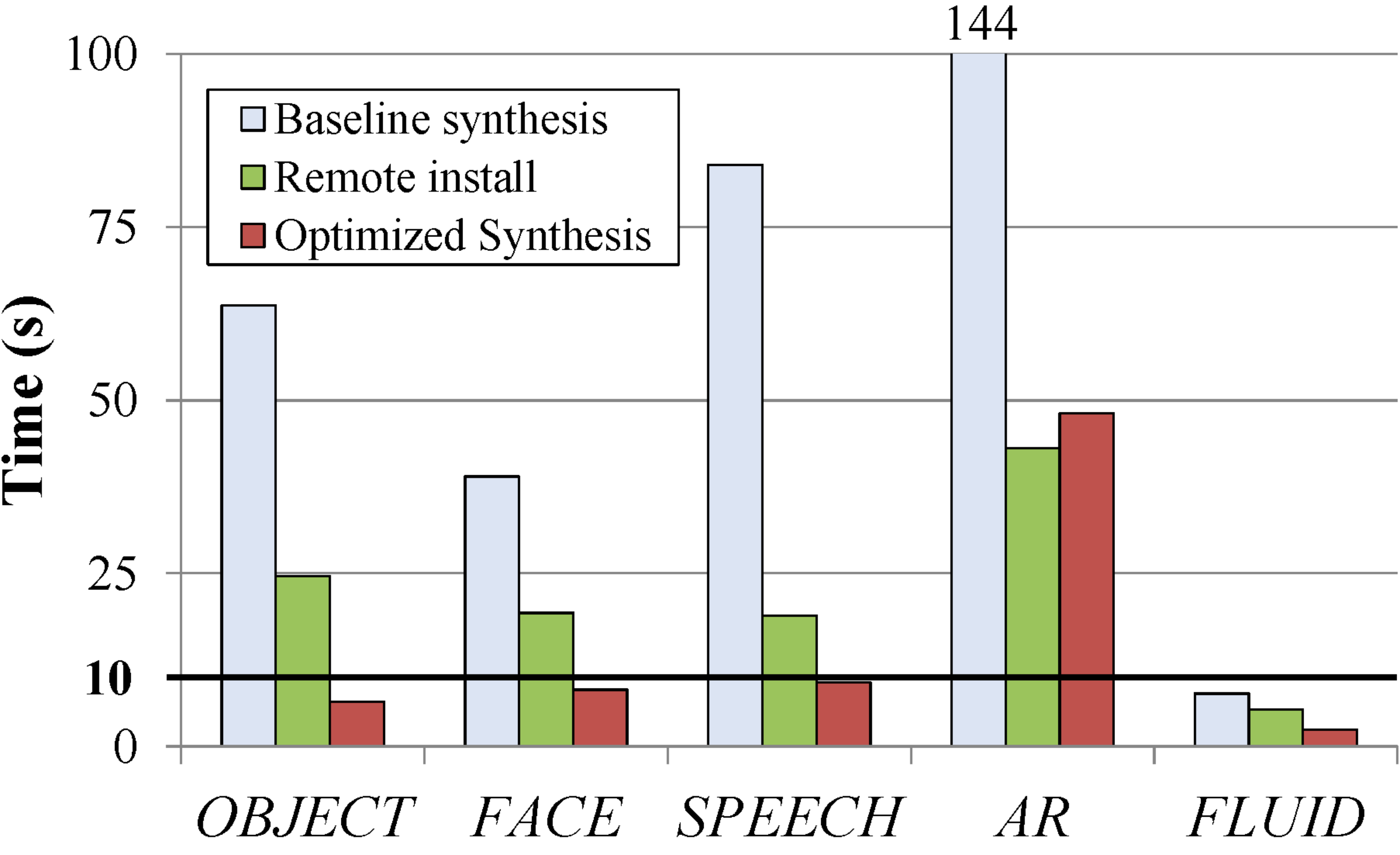


# What should a Cloudlet look like?

- Full flexibility – support any OS, app framework, partitioning methods
- Minimal management – physically install and forget model
- Decentralized and stateless
- Provisioned from cloud, user devices

  → Virtual Machines

# Rapid Provisioning of Personalized VM



# Harnessing Effortless Video Capture

## Opportunistic Sensing

EDITION: U.S. | INTERNATIONAL | MÉXICO

Sign up | Log in

CNN Justice

SEARCH

Home Video NewsPulse U.S. World Politics Justice Entertainment Tech Health Living Travel Opinion iReport Money Sports

## New Jersey family's picture catches theft in the making

By Jamie Guzzardo, CNN
August 25, 2010 9:49 a.m. EDT

Twitter Share Email Save Print

Recommend | 1600 recommendations. Sign Up to see what your friends recommend.

The picture shows a man allegedly taking John Myers' bag as his family gets photographed. Police later arrested a suspect.

**STORY HIGHLIGHTS**

- A New Jersey man snapped a photo of his family during a trip to Wisconsin
- The photo also caught the image of someone allegedly making off with the man's bag
- Among other valuables, the man's wallet and car keys were in the bag
- Wisconsin Capitol Police quickly apprehended a suspect and returned the bag

*Read more on the story on CNN affiliate WISC-TV in Madison, Wisconsin.*

**(CNN)** -- In today's technology-laden society, hearing of crimes solved or cold cases cracked with state-of-the-art tech tools has become commonplace. But for one New Jersey family all it took to catch an alleged thief was a camera and a little luck.

John Myers and his family from Bloomfield, New Jersey, were visiting Madison, Wisconsin, to attend a friend's wedding Saturday at the state Capitol.

According to Myers, the family went outside after the ceremony to

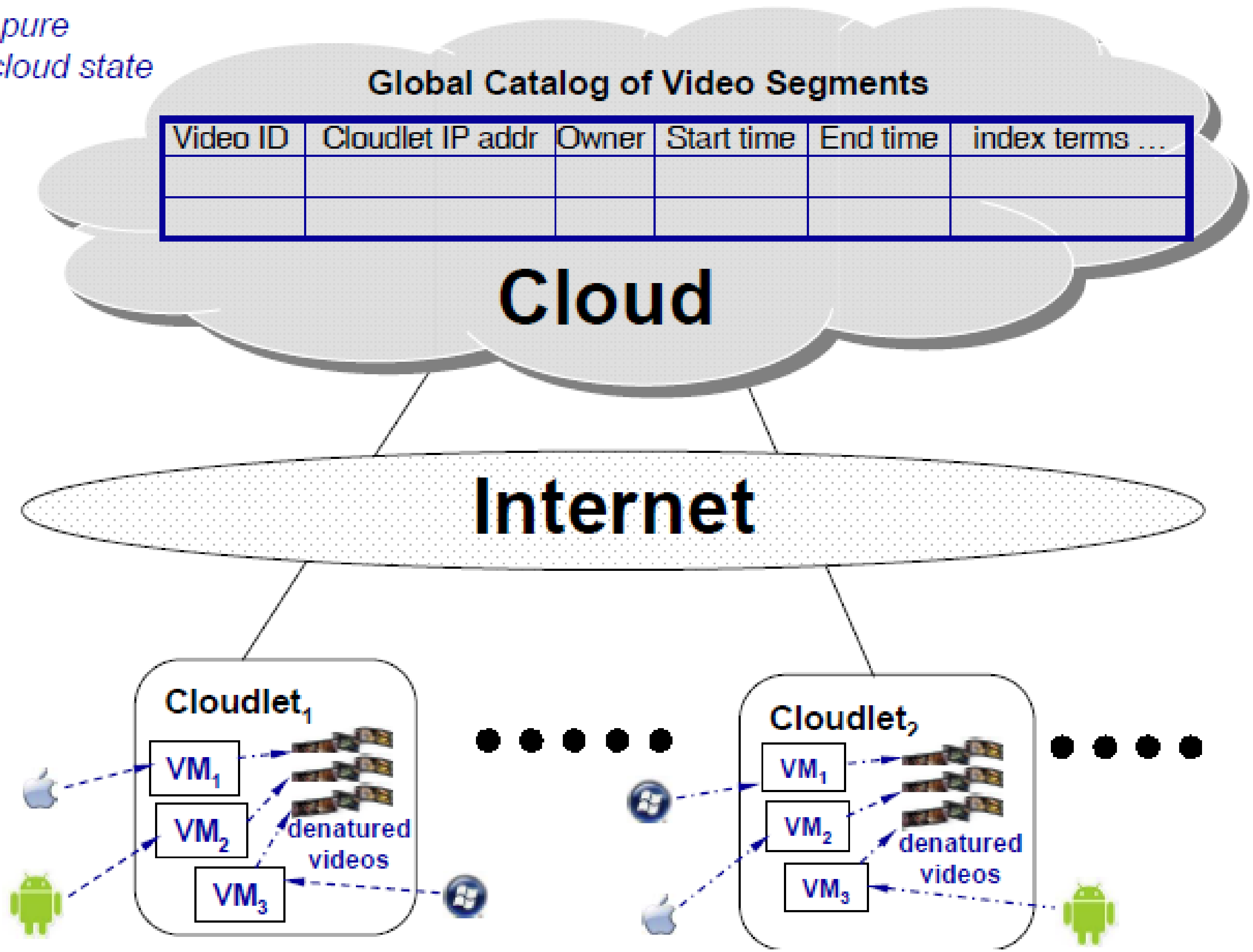


NewsPulse »

# Gigasight

*Cloudlets now contain hard state – no longer pure cache of cloud state*

**CDN in Reverse**



# Outline

- Highlights from 4 Research Pillars
  - Specialization
  - Automation
  - Big Data
  - To the Edge

- Deeper dive on Big Learning

# Big Learning Deeper Dive

Three Big Learning Frameworks @ ISTC-CC:

- Spark
- GraphLab
- Stale Synchronous Parallel

# Spark

- **Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing**, Stoica, NSDI'12, best paper
  - A restricted form of shared memory, based on coarse-grained deterministic transformations rather than fine-grained updates to shared state: expressive, efficient and fault tolerant

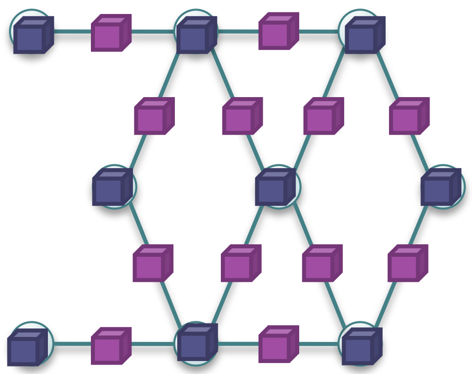
- **Discretized Streams: Fault-Tolerant Streaming Computation at Scale**, SOSP'13

Features:
- In-memory speed w/fault tolerance via logging transforms
- Bulk Synchronous

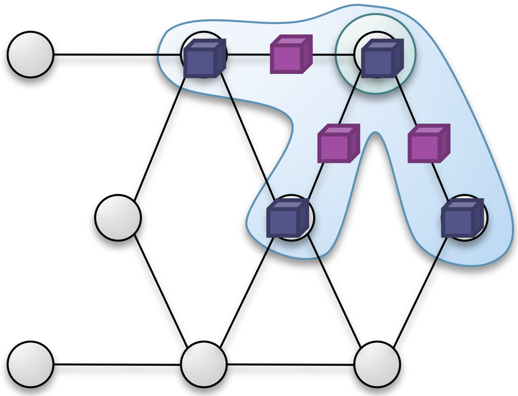
# GraphLab - 1

## Graph Parallel: "Think like a vertex"

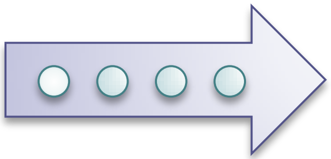
Graph Based
*Data Representation*



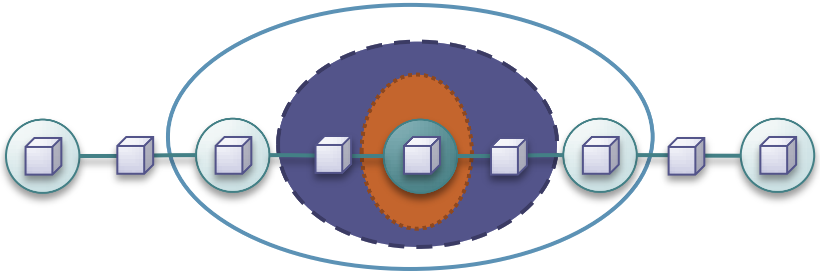
Update Functions
*User Computation*



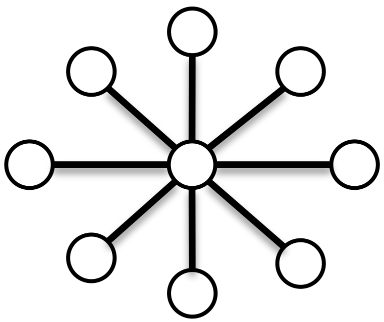
Scheduler



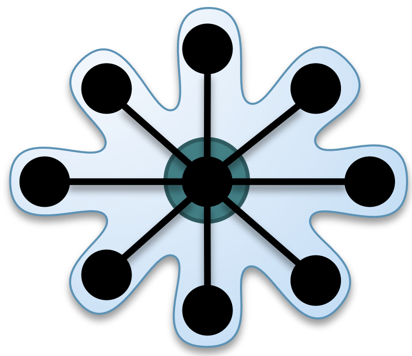
Consistency Model



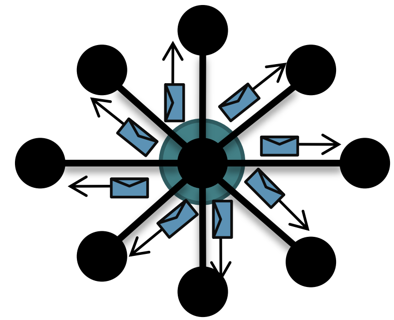
# Problem: High Degree Vertices Limit Parallelism



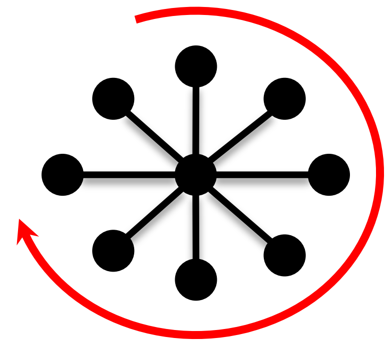
Edge information too large for single machine



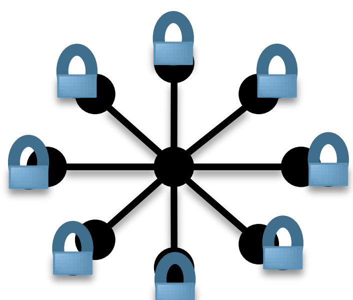
Touches a large fraction of graph (GraphLab 1)



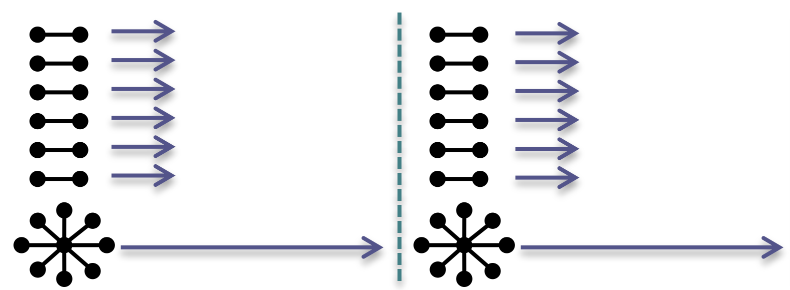
Produces many messages (Pregel)



Sequential Vertex-Updates

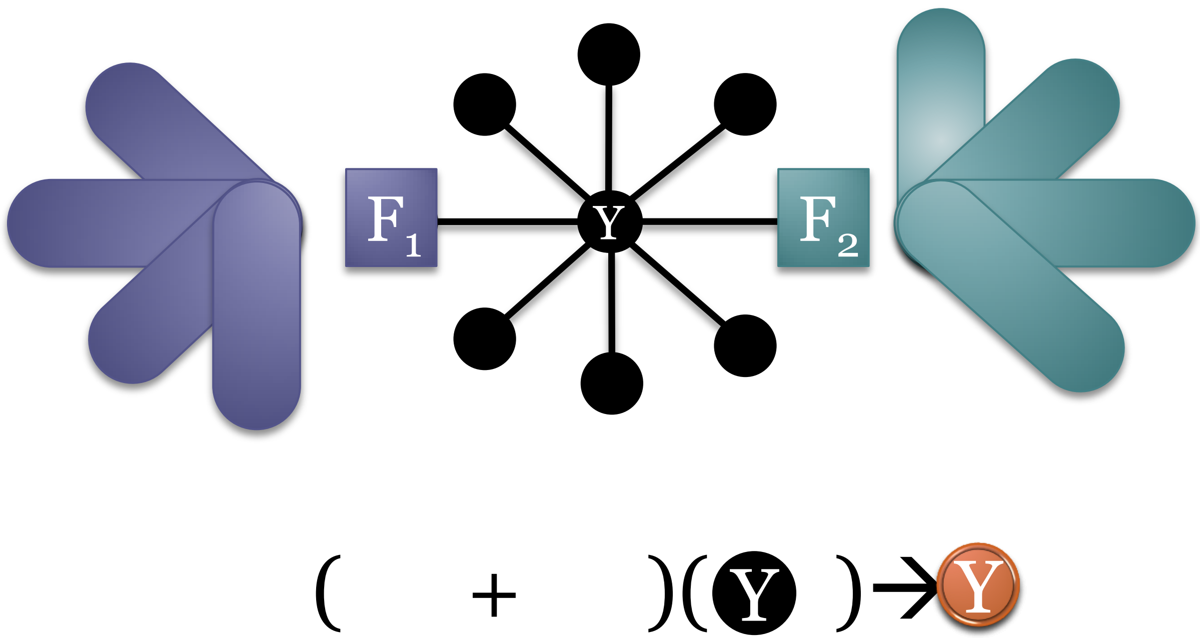


Asynchronous consistency requires heavy locking (GraphLab 1)



Synchronous consistency is prone to stragglers (Pregel)

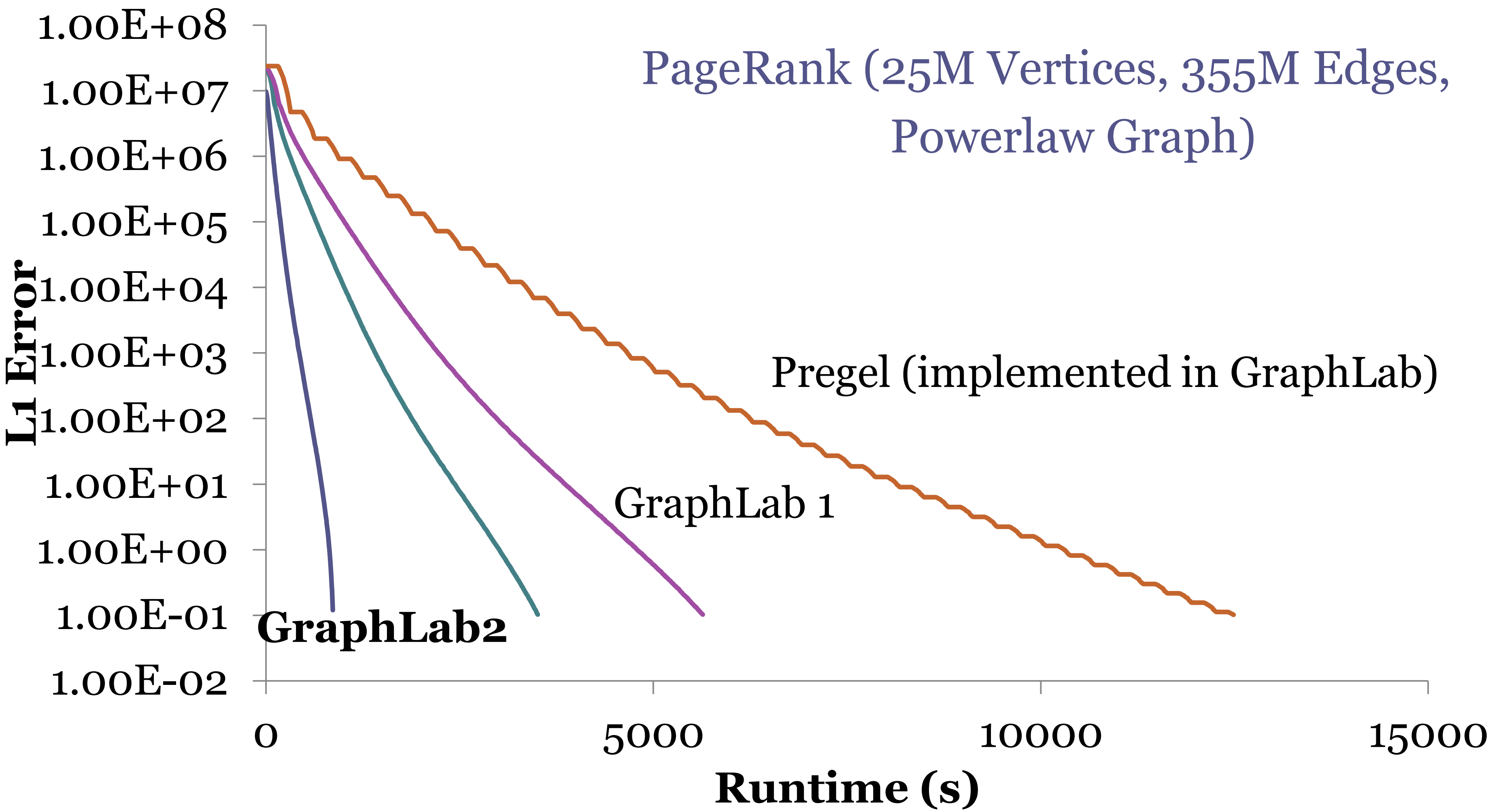
# GraphLab 2 Solution: Factorized Updates




**O(1)** data transmitted over network

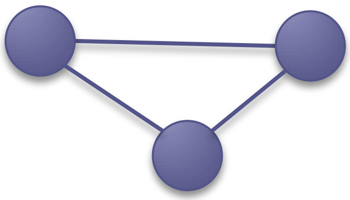
- **PowerGraph: Distributed Graph-Parallel Computation on Natural Graphs**, Guestrin, OSDI'12

# Multicore Performance



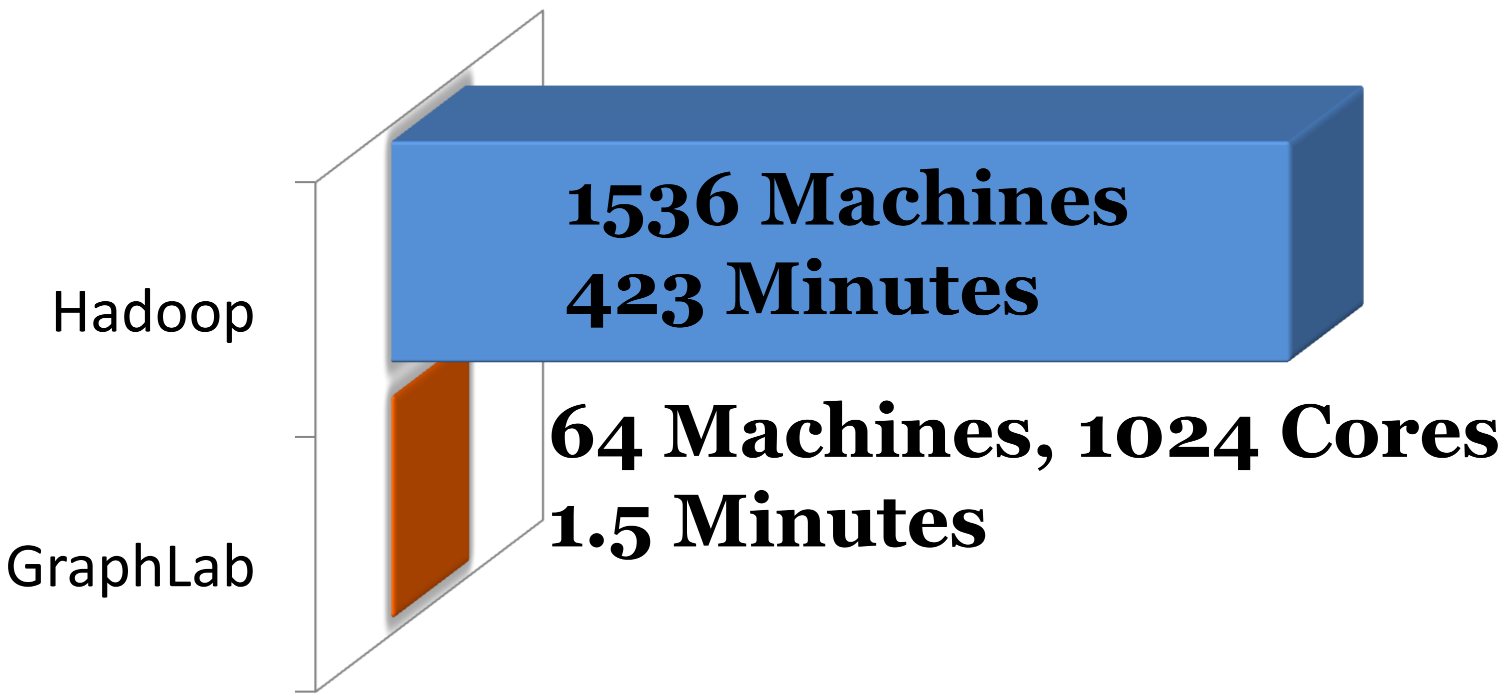

GraphLab 2 has significantly faster convergence rate

# Triangle Counting in Twitter Graph



**Total:**
**34.8 Billion Triangles**

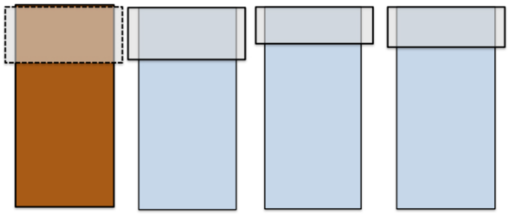
**40M Users**
**1.2B Edges**




Hadoop results from [Suri & Vassilvitskii '11]

# GraphChi – disk-based GraphLab

- Novel Parallel Sliding Windows algorithm



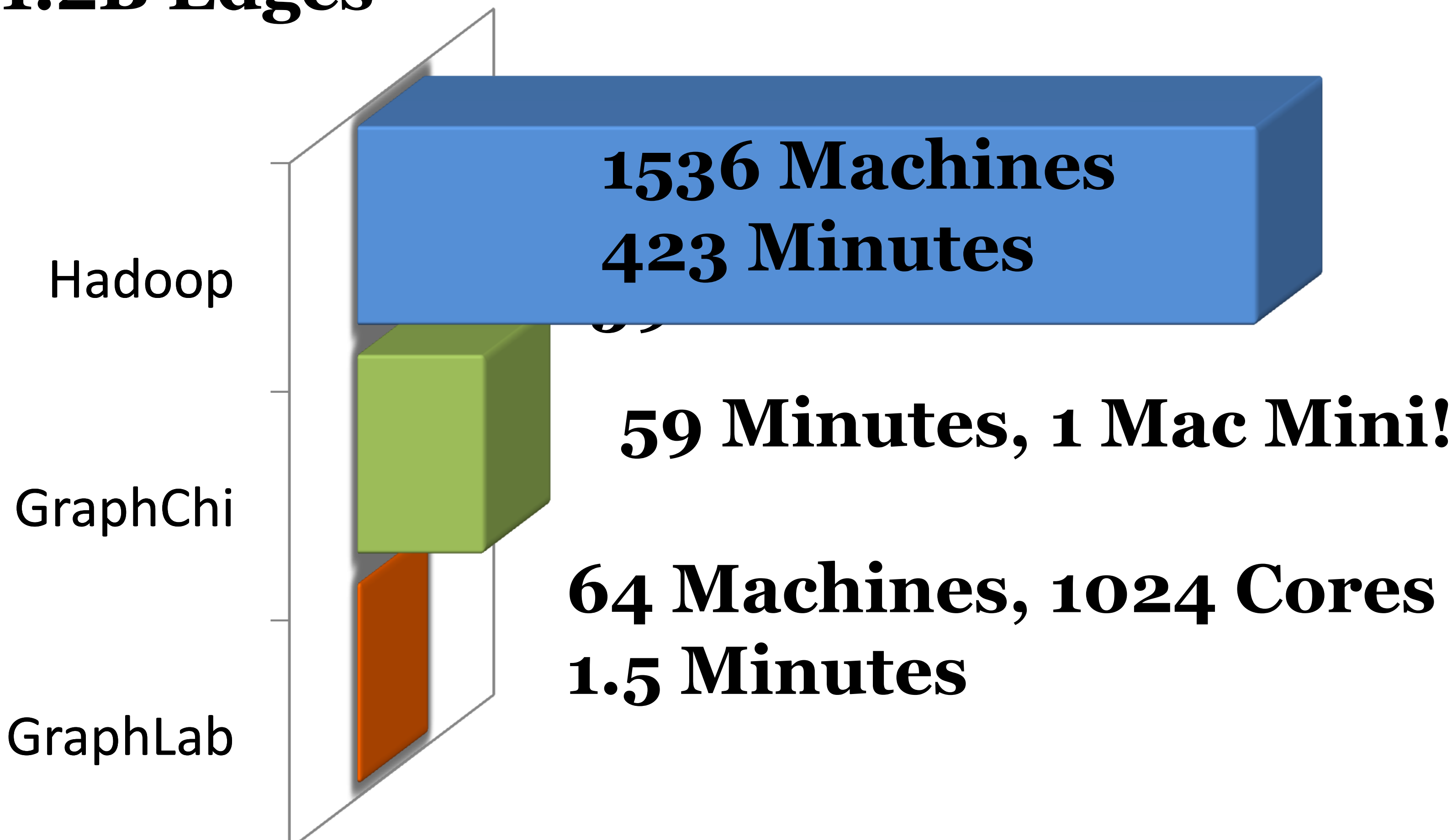

- Fast!
- Solves tasks as large as current distributed systems
- Minimizes non-sequential disk accesses
  - Efficient on *both* SSD and hard-drive
- Parallel, asynchronous execution

- **GraphChi: Large-Scale Graph Computation on Just a PC**, Guestrin, Blelloch, OSDI'12

# Triangle Counting in Twitter Graph

**40M Users**
**1.2B Edges**

**Total: 34.8 Billion Triangles**




Hadoop results from [Suri & Vassilvitskii '11]

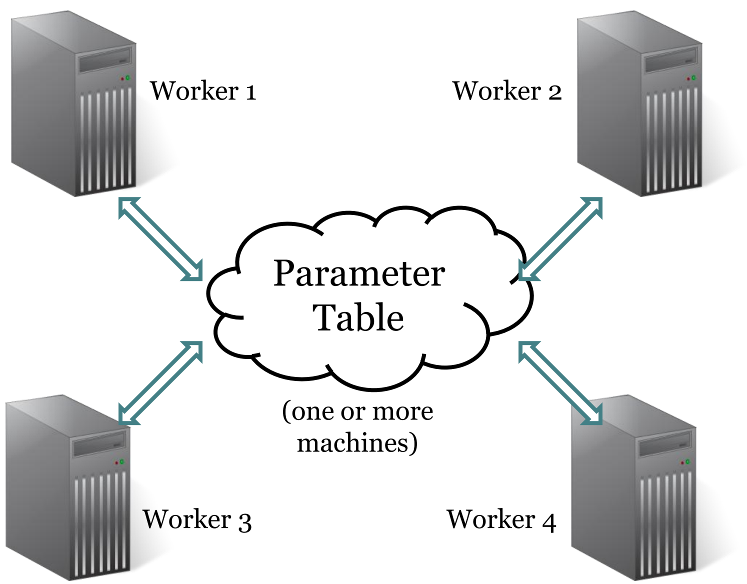
# Big Learning Deeper Dive

Three Big Learning Frameworks @ ISTC-CC:

- Spark
- GraphLab
- Stale Synchronous Parallel
  - **More Effective Distributed ML via a Stale Synchronous Parameter Server**, Ganger, G., Gibson, Xing, NIPS'13 oral

# Parameter Servers for Distributed ML

- **Provides all machines with convenient access to global model parameters**
- **Enables easy conversion of single-machine parallel ML algorithms**
  - "Distributed shared memory" programming style
  - Replace local memory access with PS access




**Single Machine Parallel**

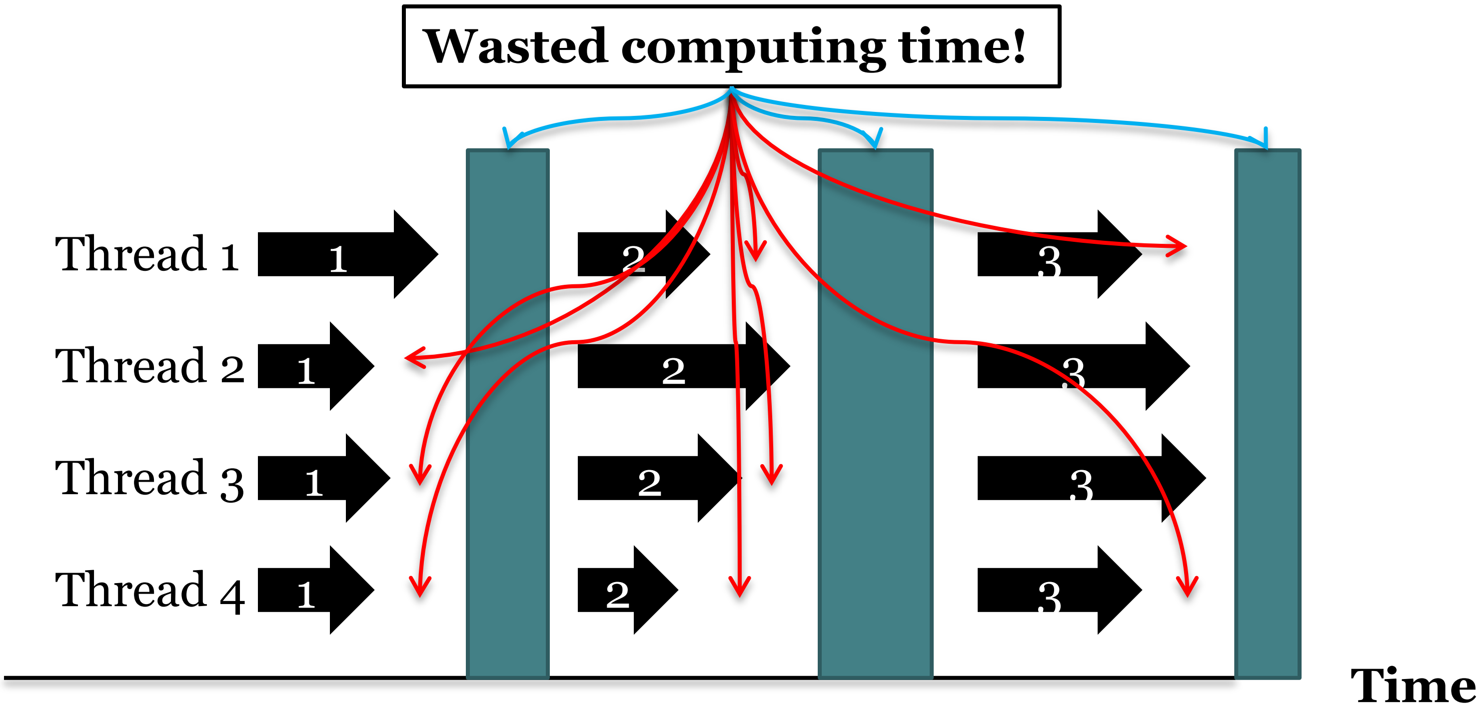
```
UpdateVar(i) {
  old = y[i]
  delta = f(old)
  y[i] += delta
}
```

**Distributed with PS**

```
UpdateVar(i) {
  old = PS.read(y,i)
  delta = f(old)
  PS.inc(y,i,delta)
}
```

† **Ahmed et al. (WSDM 2012), Power and Li (OSDI 2010)**

# The Cost of Bulk Synchrony

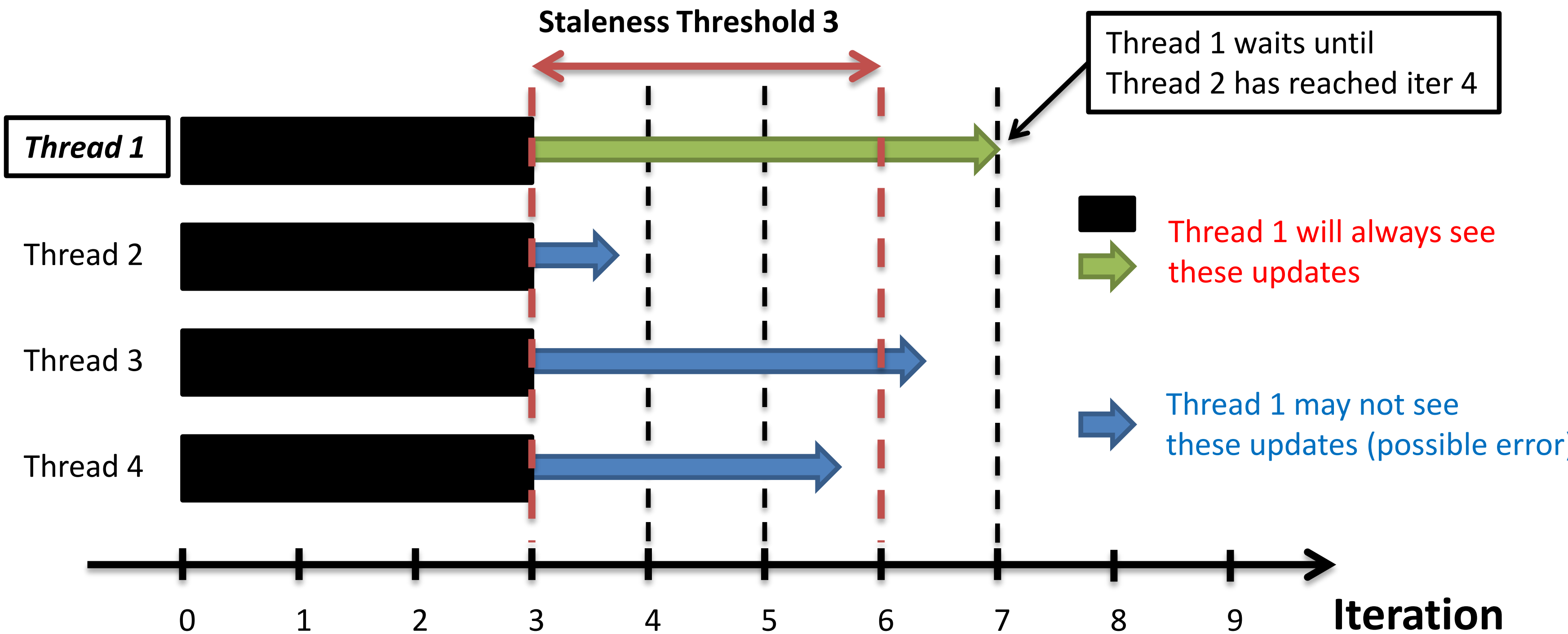



**Threads must wait for each other**
**End-of-iteration sync gets longer with larger clusters**

**Precious computing time wasted**

But: Fully asynchronous => No algorithm convergence guarantees

# Stale Synchronous Parallel




**Allow threads to *usually* run at own pace**

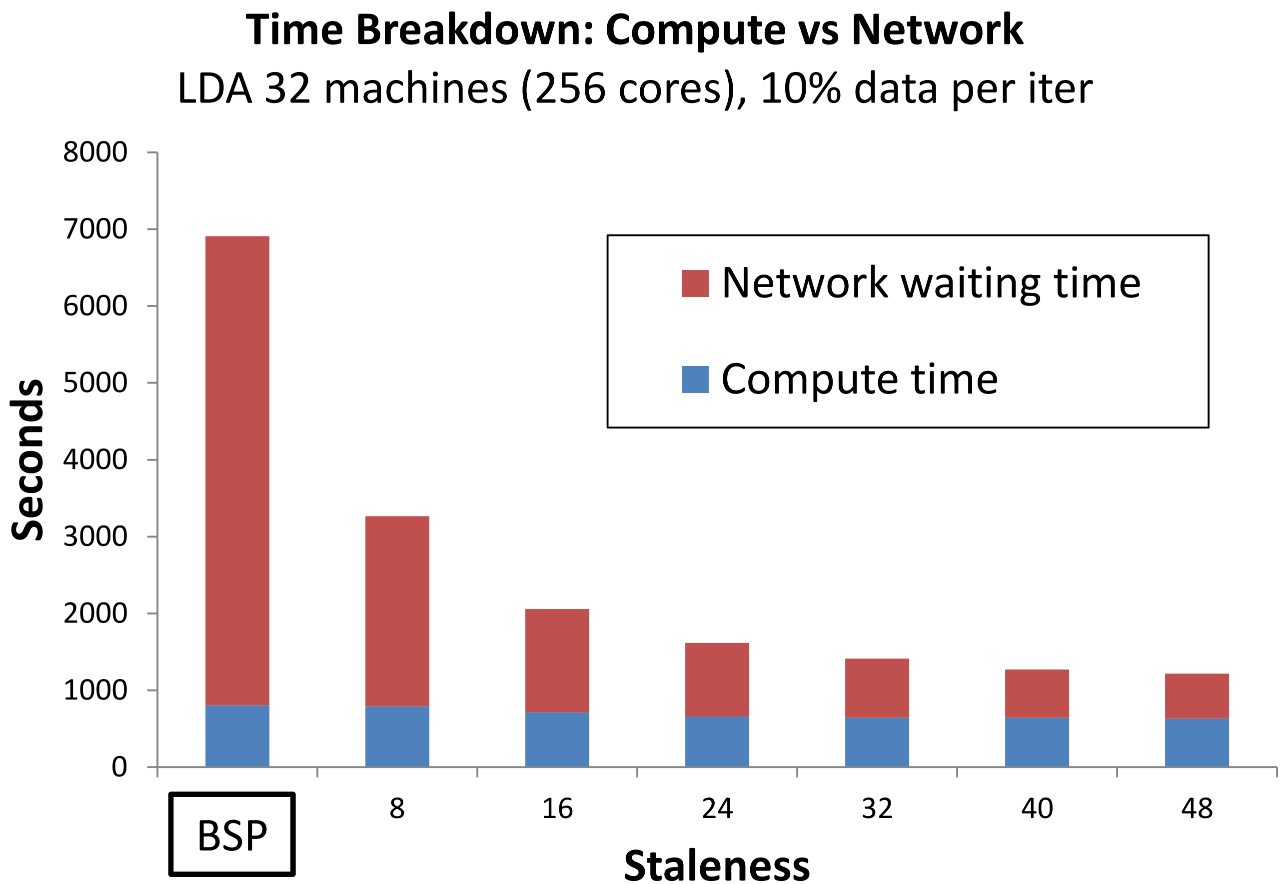
Fastest/slowest threads not allowed to drift >S iterations apart

Protocol: check cache first; if too old, get latest version from network

Consequence: fast threads must check network every iteration

**Slow threads check only every S iterations** – fewer network accesses, so catch up!

# SSP uses networks efficiently

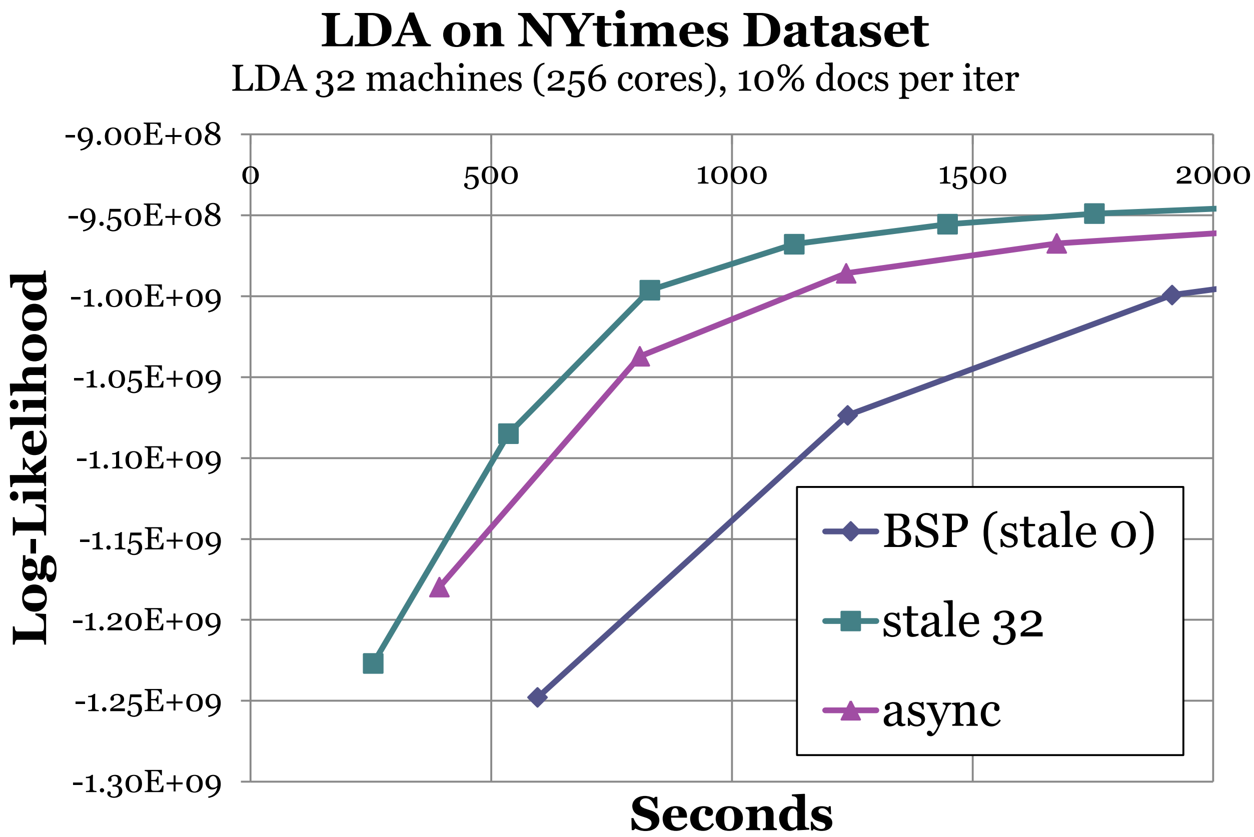



Network communication is a huge bottleneck with many machines

**SSP balances network and compute time**

# SSP vs BSP and Async

NYtimes data
N = 100M tokens
K = 100 topics
V = 100K terms




BSP has strong convergence guarantees but is slow
Asynchronous is fast but has weak convergence guarantees
**SSP is fast and has strong convergence guarantees**

# ISTC-CC: Research Projects

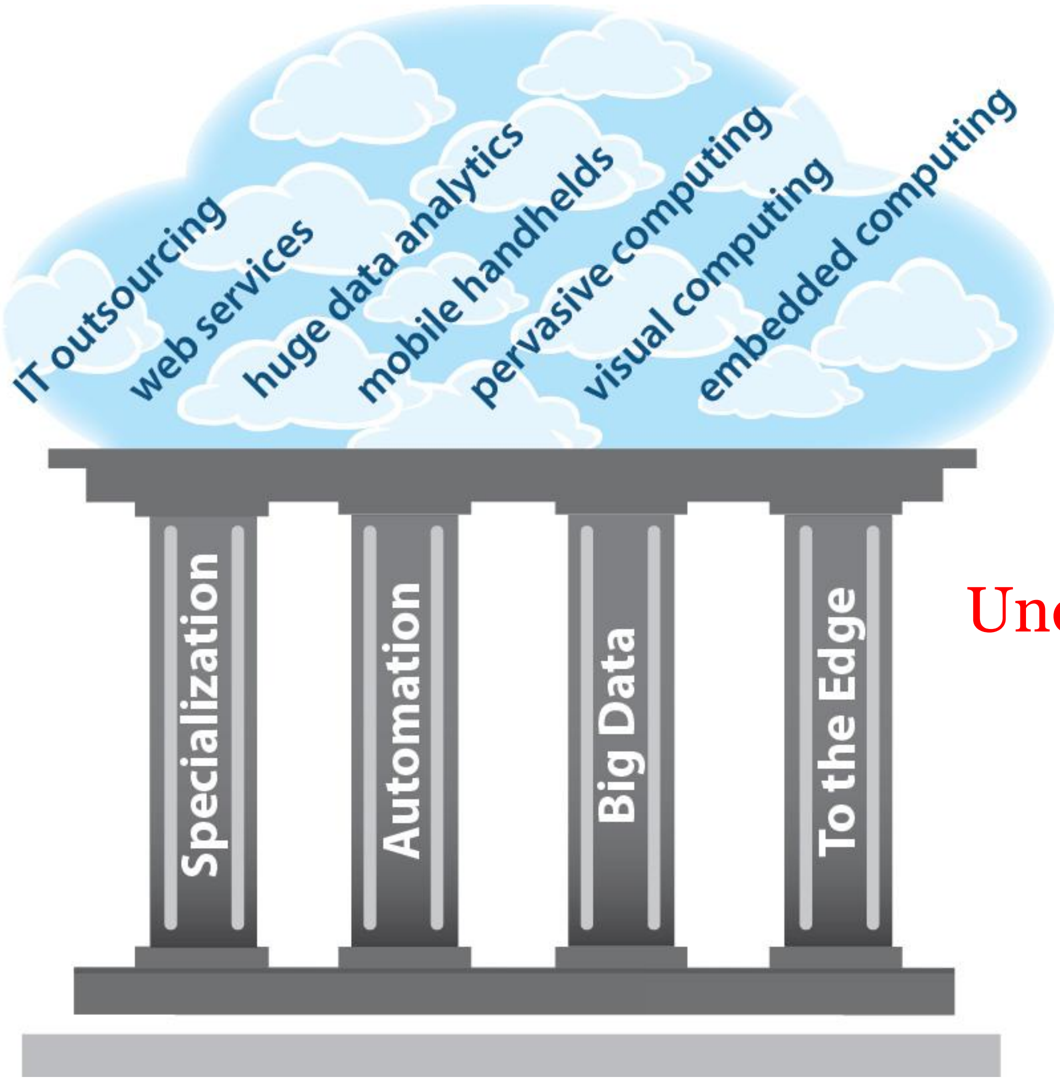
| | Project | Personnel |
|---|---|---|
| **S1** | Specialized Platforms of Wimpy Nodes | **Andersen[C], Schwan[G], Freedman[P], Kaminsky[I], Kozuch[I], Pillai[I]** |
| **S2** | Specialized Platforms of Heterogeneous Many-Cores | **Mowry[C], Mutlu[C], Gavrilovska[G], Schwan[G], Yalamanchili[G], Martonosi[P], Gibbons[I], Kozuch[I]** |
| **A1** | Resource Scheduling for Heterogeneous Cloud Infrastructures | **Joseph[B], Katz[B], Stoica[B], Ganger[C], Harchol-Balter[C], Kozuch[I]** |
| **A2** | Problem Diagnosis and Mitigation | **Ganger[C], Narasimhan[C], Eisenhauer[G], Liu[G], Schwan[G], Wolf[G]** |
| **B1** | Big Learning Systems | **Stoica[B], Andersen[C], Blelloch[C], Ganger[C], Gibson[C], Smola[C], Xing[C], Guestrin[W], Gibbons[I]** |
| **B2** | Big Data Storage | **Andersen[C], Ganger[C], Gibson[C], Xing[C], Pu[G], Schwan[G]** |
| **E1** | Cloud-Assisted Mobile Client Computations | **Satya[C], Siewiorek[C], Gavrilovska[G], Liu[G], Schwan[G], Martonosi[P], Pillai[I]** |
| **E2** | Geographically Distributed Data Storage | **Andersen[C], Satya[C], Siewiorek[C], Freedman[P], Kaminsky[I], Pillai[I]** |

# Open Source Code Releases in Year 2

- GraphBuilder 1.0 released open source in Jun'13
- GraphLab 2.2 released open source in Jul'13
- Spark 0.8 release Sep'13 – Apache incubator
- Mesos 0.14 released Oct'13 – Apache
- Other open source releases on github include: Eiger, EPaxos, Parrot, Cloudlet OpenStack++, CuckooFilter, RankSelect, MemC3, NVMalloc, etc.

Open Source page: www.istc-cc.cmu.edu/research/ossr/

# Intel Science & Technology Center for Cloud Computing




Underlying Infrastructure
enabling the future
of cloud computing

www.istc-cc.cmu.edu

# Slide Credits

A number of these slides were adapted from slides created by the following ISTC-CC Faculty:

- Dave Andersen, Greg Ganger, Garth Gibson, Carlos Guestrin, Onur Mutlu, Priya Narasimhan, Babu Pillai, M. Satyanarayanan, and Eric Xing

...and their students