



The Future of Cloud Computing

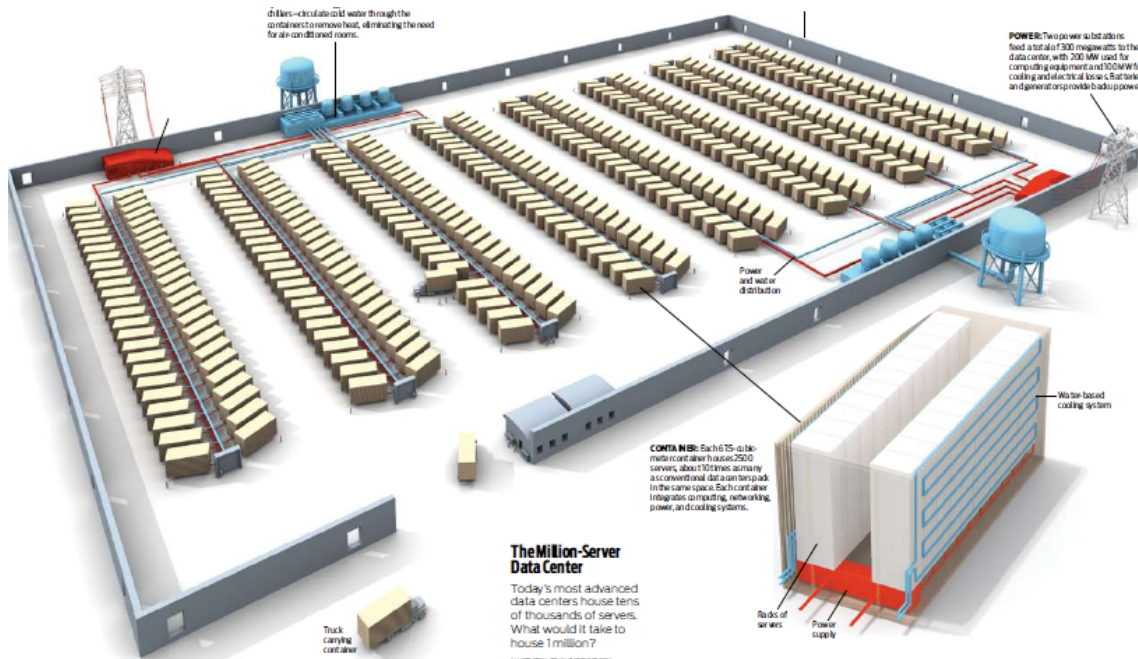
Michael Franklin, UC Berkeley
Reliable Adaptive Distributed Systems Lab





Datacenter is the new Server

Utility computing: enabling innovation in new services without first building & capitalizing a large company.



The Million-Server Data Center

Today's most advanced data centers house tens of thousands of servers. What would it take to house 1 million?

ILLUSTRATION: WIKI/ISTOCKPHOTO.COM



RAD Lab 5-year Mission

Enable 1 person to develop, deploy, operate next-generation Internet application

- Key enabling technology: Statistical machine learning
 - debugging, power management, performance prediction, ...
- Highly interdisciplinary faculty & students
 - PI's: Fox/Katz/Patterson (systems/networks), Jordan (machine learning), Stoica (networks & P2P), Joseph (systems/security), Franklin (databases)
 - 2 postdocs, ~30 PhD students, ~10 undergrads

The multi-colored Google logo.The Microsoft logo in a bold, black, sans-serif font.The Sun logo, featuring a blue square with a white geometric pattern and the word "Sun" in a blue serif font.The Amazon Web Services logo, featuring a yellow smile arrow and the text "amazon web services".The Cisco logo, featuring a stylized bridge icon and the word "CISCO" in a bold, black, sans-serif font.The Cloudera logo, featuring a blue and white geometric pattern and the word "cloudera" in a blue, lowercase, sans-serif font.The eBay logo, featuring the word "eBay" in a stylized, multi-colored font.The Facebook logo, featuring the word "facebook" in a white, lowercase, sans-serif font on a blue rectangular background.The Fujitsu logo, featuring a stylized red and white geometric pattern and the word "FUJITSU" in a red, uppercase, sans-serif font.The HP logo, featuring the letters "hp" in a white, lowercase, sans-serif font inside a blue circle, with the word "invent" in a blue, lowercase, sans-serif font below it.The Intel logo, featuring the word "intel" in a blue, lowercase, sans-serif font inside a blue circle.The NetApp logo, featuring a blue square with a white geometric pattern and the word "NetApp" in a blue, uppercase, sans-serif font below it.The SAP logo, featuring the word "SAP" in a white, uppercase, sans-serif font inside a blue square.The VMware logo, featuring the word "vmware" in a blue, lowercase, sans-serif font.The Yahoo! Research logo, featuring the word "YAHOO!" in a purple, uppercase, sans-serif font with an exclamation point, and the word "RESEARCH" in a blue, uppercase, sans-serif font below it.



Above the Clouds: A Berkeley View of Cloud Computing

abovetheclouds.cs.berkeley.edu

- 2/09 White paper by RAD Lab PI's and students
 - Clarify terminology around Cloud Computing
 - Quantify comparison with conventional computing
 - Identify Cloud Computing challenges & opportunities
 - stimulate discussion on *what's really new*
- ~60K downloads; >170 citations;
 - “Circulated to CxOs” of major IT firms
 - “profound effect” on datacenter strategy
 - Short version to appear in March 2010 CACM

What's new in the cloud?

- Not-so-new: Software as a Service (SaaS)
 - Basic idea predates MULTICS
 - Software hosted in the infrastructure
 - Recently: “[HW, Infrastructure, Platform] as a service” ??
- **New:** *pay-as-you-go utility computing*
 - Illusion of infinite resources on demand
 - Fine-grained billing: release == don't pay
 - *Economies of Scale:*
 - *DC resources 5-7x cheaper than medium-sized (100s servers)*
 - *Statistical multiplexing enables increased utilization*
 - *Public (utility) vs. private clouds*



Not Just CapEx vs. OpEx

- “Cost associativity”: 1,000 CPUs for 1 hour same price as 1 CPU for 1,000 hours (@\$0.10/hour)
 - RAD Lab graduate students demonstrate improved Hadoop (batch job) scheduler—on 1,000 servers
- *Risk Transfer* for SaaS startups
 - *Animoto* traffic doubled every 12 hours for 3 days when released as Facebook plug-in
 - Surged from 50 to >3500 servers. ...***then back down***
- Gets IT gatekeepers out of the way
 - not unlike the PC revolution

Cloud Usage Types

- Interactive
 - Webapps, E-Commerce, Media hosting
- Analytic
 - Business Intelligence, High-Performance Computing
- Infrastructure
 - Backup & Storage, Content Delivery

RAD Lab Prototype: System Components




log
mining

Perf
Stats

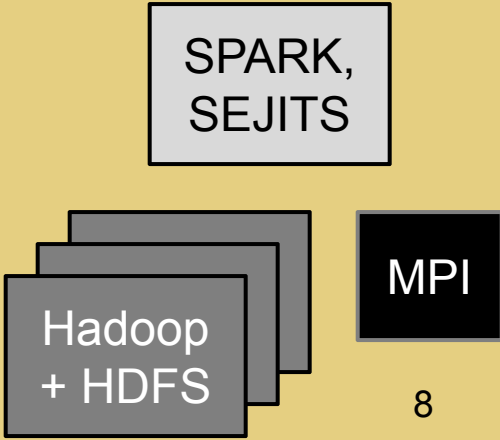
Xtrace + Chukwa
(monitoring)

WebApp/RoR
PIQL – Query Language

SCADS - ScalableStorage



NEXUS – Cloud OS



- **Recurring theme:** cutting-edge Statistical Machine Learning (SML) works where simpler methods have failed
 - Predict performance of complex software system when demand is scaled up
 - Automatically add/drop servers to fit demand, without violating Service Level Agreement (SLA)
 - Distill millions of lines of log messages into an operator-friendly “decision tree” that pinpoints “unusual” incidents/conditions

**See posters and meet researchers at the
RADLab Open House this afternoon.**



So, where are things going?

Continuous Improvement of Input Devices



GarageBand '08



Ubiquitous Connectivity





amazon **mechanical turk**
beta Artificial Intelligence

Make Money by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)

As a Mechanical Turk Worker you:

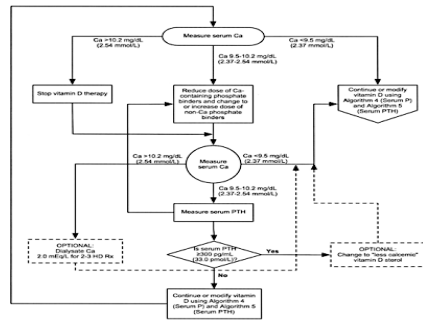
- Can work from home
- Choose your own work hours
- Get paid for doing good work



The Scalability Dilemma

- State-of-the Art Machine Learning techniques do not scale to large data sets.
- Data Analytics frameworks can't handle lots of incomplete, heterogeneous, dirty data.
- Processing architectures struggle with increasing diversity of programming models and job types.
- Adding people to a late project makes it later.

Exactly Opposite of what we Expect and Need



Adaptive/Active Machine Learning and Analytics



Massive and Diverse Data

CrowdSourcing

Cloud Computing

Participatory Culture - Direct



Web 2.0

Name Generator

[Dot-o-mator](#)

[Domain Naming Tips](#)

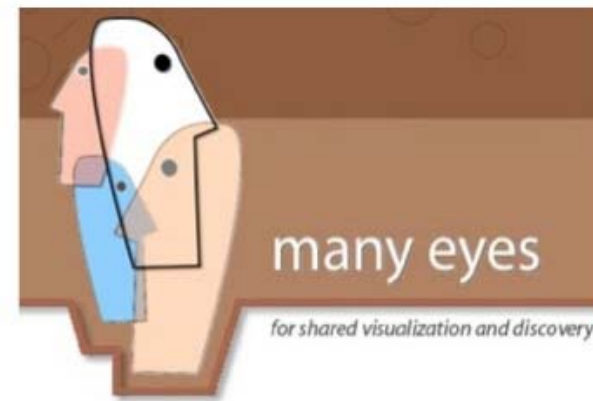
Need a name for your Web 2.0 company? Try this handy name generator.

No, I don't promise that any of these names are actually available.

[Generate Name!](#)

Blogzoom

[Check Availability!](#)



Participatory Culture – “Indirect”

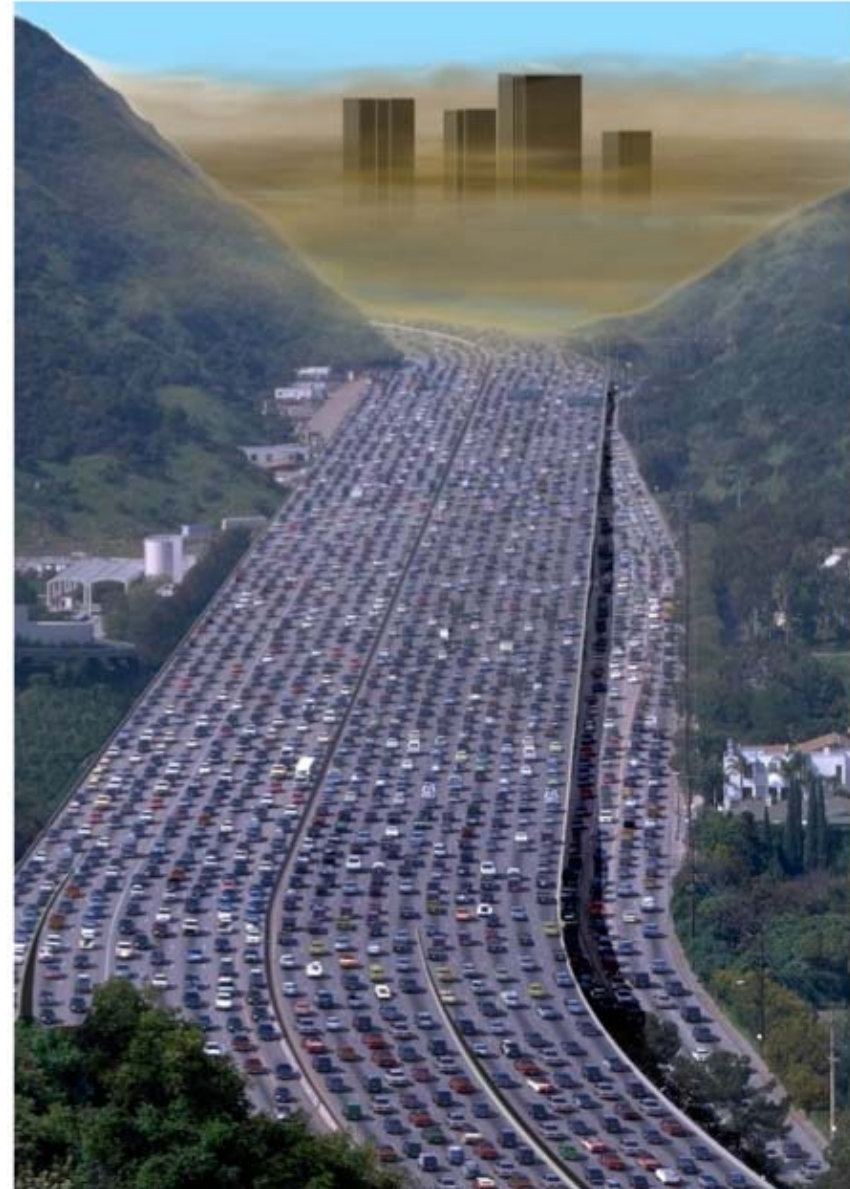
John Murrell: GM SV 9/17/09

...every time we use a Google app or service, we are working on behalf of the search sovereign, creating more content for it to index and monetize or teaching it something potentially useful about our desires, intentions and behavior.





Use Case: Privacy-enhanced Traffic Crowdsourcing w/ Alex Bayen CEE/CITRIS



Use Case: Urban Planning μ -simulation w/Paul Waddell, Environmental Design




Use Case: Computational Journalism w/UCB J-School

EveryBlock *A news feed for your block*

Track and discuss what's new in your neighborhood.

Choose a city...

- BETA **Atlanta**
- Boston**
- Charlotte**
- Chicago**
- BETA **Dallas**
- BETA **Detroit**
- BETA **Houston**
- Los Angeles**
- Miami**
- New York**
- Philadelphia**
- San Francisco**
- San Jose**
- Seattle**
- Washington, DC**



The image shows three overlapping screenshots of the EveryBlock platform. The top screenshot is the New York City 'Restaurant inspections' page, featuring a bar chart for the period Dec 13, 2008 - Jan 12, 2009, showing 2,573 inspections. The middle screenshot shows the Chicago interface for the 1050-1098 block, displaying a 'Latest news' timeline and a 'Filter this data' section with options for location, inspection date, and violation type. The bottom screenshot shows the mobile app interface for Chicago, displaying a 'Nearby' filter and a news item for 'Monday, April 13, 2009' with a 'Five things to do today' section.

Clouds and Crowds

	Interactive Cloud	Analytic Cloud	People Cloud
Data Acquisition	Transactional systems Data entry	... + Sensors (physical & software)	... + Web 2.0
Computation	Get and Put	Map Reduce Parallel DBMS Stream Processing	... + Collaborative Structures (e.g., Mechanical Turk, Intelligence Markets)
Data Model	Records	Numbers, Media	... + Text, Media, Natural Language
Response Time	Seconds	Hours/Days	... +Continuous

The Future Cloud will be a Hybrid of These.

Technical Challenges

- BIG DATA Machine Learning & Analytics
- Text analytics
- Data Integration & Management
- Systems and Programming Frameworks
- Collaboration structures & Visualization
- **Hybrid Cloud/Crowd** scheduling, resource management, ...

AMP Lab: The Next Generation

Enable many people to collaborate to collect, generate, clean, make sense of and utilize lots of data.

- Approach: An end-to-end view of the entire stack from data visualization down to cluster & multicore support.
- Highly interdisciplinary faculty & students
- Developing a five-year plan, will dovetail with RADLab completion
- Sponsors:

- RAD Lab Open House, 465 Soda Hall
 - posters, research, students, faculty
 - discussion of AMP Lab planning
- abovetheclouds.cs.berkeley.edu
 - Paper, executive summary, slides
 - “Above the Clouds” blog
 - Impromptu video interview with authors
- franklin@cs.berkeley.edu