

Building Robust Applications in Cloud

Yuanyuan (YY) Zhou

Dept. of Computer Science
University of California, San Diego

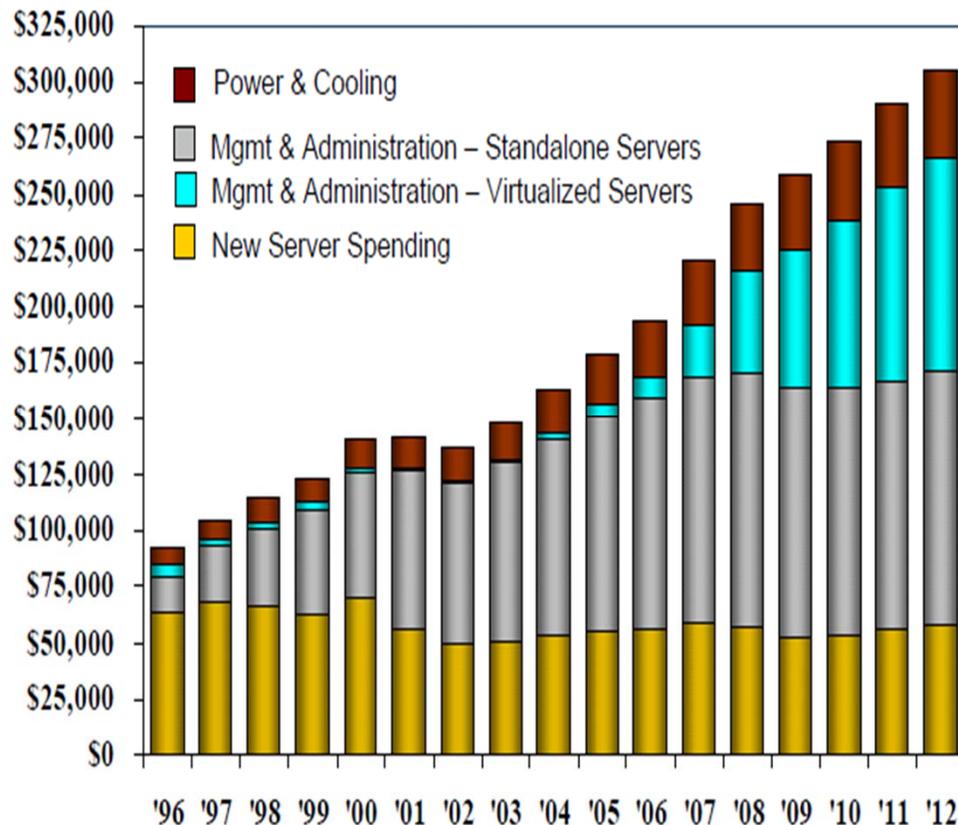
yyzhou@cs.ucsd.edu

<http://www.cs.ucsd.edu/~yyzhou>



Benefits of a Cloud Environment

Worldwide IT spending – server related



Source: March 2009 IDC Directions

Reduced Operating costs

- Reduce Equipment Expenses via consolidation
- Reduce Labor and Administration cost by increasing server administration ratios from 20-35:1 to over 400:1 in a cloud farm environment

Improved Service Levels

- Reduce Provisioning Time & Cost from hours /days to minutes.

Reduce Capital Expenditures

- Significantly reduce need to add new equipment (servers, storage, and network)
- Reposition costs from capital expense to operating expense

Commercial Clouds

Cloud/ Service	Amazon	Microsoft Azure	Google (and Apache)
Data	S3, EBS, SimpleDB	Blob, Table, SQL Services	GFS, BigTable
Computing	EC2, Elastic Map Reduce (runs Hadoop)	Compute Service	MapReduce (not public, but Hadoop)
Service Hosting	None?	Web Hosting Service	AppEngine / AppDrop



Who is Using Clouds?



Example 1: **Mogulus**

- A live broadcast platform on the internet
 - Producers can use its studio application to create LIVE, on-demand internet television to broadcast on the web
- Entirely hosed on cloud
- On election night, 87000 videos @500kbps = 43.5Gbps
- <http://www.mogulus.com>



Example: New York Times

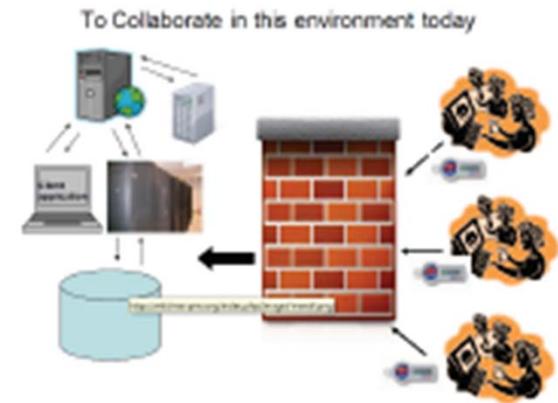
- **Timesmachine**: a news archive of the NY Times available in pdf over the Internet for subscribers
- Entirely hosted on cloud
- Several terabits of data
 - Business owners got to data up on cloud for **\$50 over one weekend**

Welcome to
Times Machine
Browse 70 years of New York Times archives



Example: Eli Lilly

- 10th largest pharmaceutical company in the world
- Moved **entire R&D environment** to cloud
- Results:
 - Time to deliver new services greatly reduced
 - New server: 7.5 weeks → 3min
 - 64 node linux cluster: 12 weeks → 5min



But There are Issues & Concerns...

GLOBAL SERVICE DELIVERY



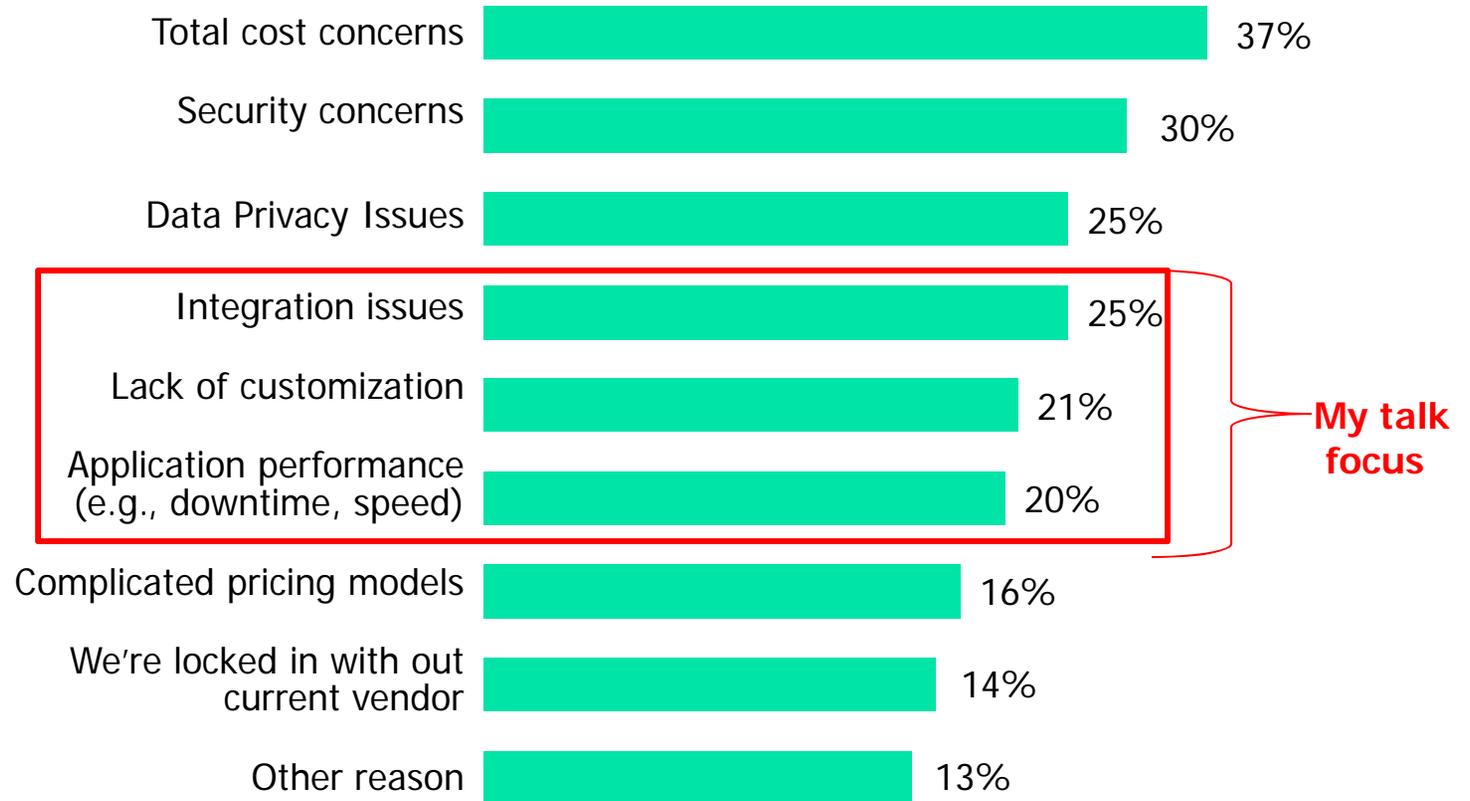
OPTIMIZED SERVICE DELIVERY



IT SERVICE MANAGEMENT

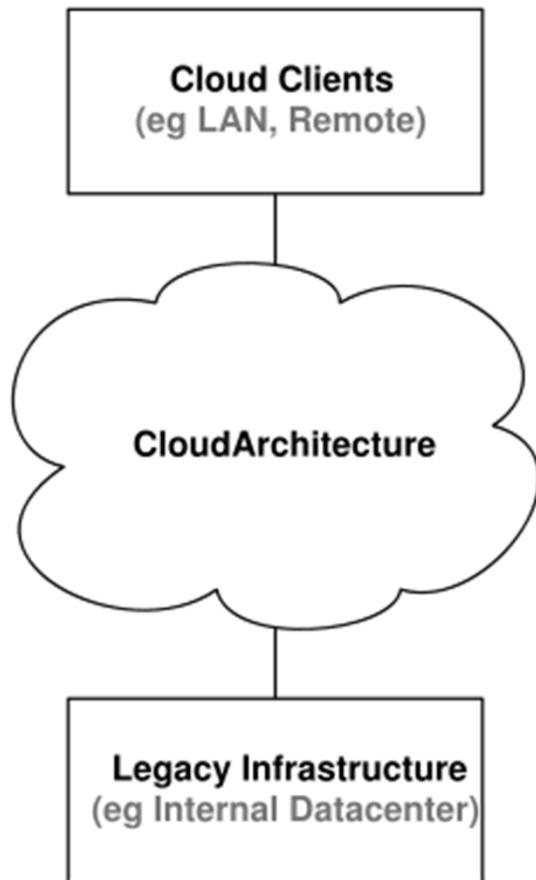
FORRESTER

* Q4 Survey of Potential Cloud Computing buyers



Many barriers to adoption, such as security, Data Privacy, DC Integration Issues, and SLA's remain. This is especially true for mission critical data center environments

Virtualization Makes it Easier and Harder



- **Easier:**
 - No need to consider hardware, other platform components
 - Scale on demand
- **Harder:**
 - What to expect?
 - Less control and transparency
- **Question: is virtualization 100%?**
 - Performance virtualizable?
 - Reliability virtualizable?



Complexity Hurts Robustness

- Traditional Distributed System Challenge
 - Unreliable Hardware
 - Component interaction
- Transparency in clouds(Lack of Control)
- Elasticity : dynamically grow and shrink
 - How to test it?
- Interference by other applications



A Real World Bug Example

- **News:** Microsoft confirmed that a bug affecting its Windows Azure operating system rendered some applications running in the Cloud inaccessible between 1:15pm and 9:00pm in Apr 3rd 2009
- **Reason:** due to a bug in the periodic application status checking module of the fabric controller software, some applications become inaccessible
- **Question:** When something fails, who is at fault?
 - Application itself? or the cloud infrastructure for which you don't have any control and accessibility?



Another Example

The cloud bites back: Google bug shared private Google Docs data

By Andrew Nusca | March 10, 2009, 8:20am PDT

Summary

Google has confirmed that a software bug in its Google Docs online application service exposed documents

Google has confirmed that a software bug in its **Google Docs** online application service exposed documents thought to be privately stored.

The problem was fixed by the weekend, and is believed to have affected only half a percent of the digital documents at a Google Docs service that provides text-handling programs as services on the Internet.



So what's a big deal?

---Bug's Cost

- 44% of respondents in a survey by *ElectricCloud* estimated that a bug costs
 - an average of \$250,000 in lost revenue and
 - 20 developer-hours to correct it



What are Needed?

- Training developers to think cloud
- Cloud testing framework
- Application Monitoring
- Diagnostic Support
- Portability to other clouds



Training Developers to Think Cloud

- Difference from traditional development
 - Statelessness and failures are givens in the cloud
 - Prepare for failures
 - Lack of transparency and control (virtual vs. physical)
 - Log more diagnostic information
 - Network, I/O contention with other applications
 - Expect interference
 - Less individualized platform support for reliability and robustness
 - Check what is available
 - Elasticity
 - Design and test it



Testing, testing, testing...

Survey Finds 58% of Software Bugs Result from Test Infrastructure and Process, Not Design Defects

Developers Prefer Taxes to Dealing with Software Testing

Sunnyvale, Calif. — June 2, 2010 Electric Cloud®, the leading provider of software production management (SPM) solutions, today released the results of a survey conducted in partnership with Osterman Research showing that the majority of software bugs are attributed to poor testing procedures or infrastructure limitations rather than design problems. Additionally, the software test process is generally considered an unpleasant process, with software development professionals rating the use of their companies' test systems more painful than preparing taxes.

Fifty-eight percent of respondents pointed to problems in the testing process or infrastructure as the cause of their last major bug found in delivered or deployed software, not design defects.



How to test for cloud?

- **Cloud providers need to provide a similar environment for testing**
 - Same resource allocation policy
 - Same load balancing policy
 - Similar infrastructure
- **Testing for elasticity**
 - Creating dynamic workloads varying from heavy to light
- **Distributed testing**
 - An old topic, but still a problem
 - Used to be a problem only for an elite few



Monitoring Cloud Applications

- Monitoring resource utilization
 - Useful to deciding usages
 - Useful to detecting abnormal behavior
- Event monitoring and alert
 - Identify failures and possible intrusions
- A lot of acquisition activities lately in industry



Diagnosing Application on Clouds

- The biggest challenge
 - Fault localization: who is at fault?
 - Multi-vendor diagnosis: network, system, storage, database, application, etc
- May require:
 - Multiple vendors to work together to gather diagnostic information or auto-diagnosing
 - Need support (log collection, search, dependency analysis) to help diagnosing faults in applications



Conclusion: Path to Cloud

Maturity Level	Provider for this service exists	Demonstrable resilient architecture	SLAs are available	Option to run the service in-house	Evidence of portability between service providers	3rd party assurance and monitoring of services	A marketplace of providers exist with easy switching between providers
I need a lot of convincing.	Yes						
Caveat emptor.	Yes	Yes	Yes				
Looks promising.	Many	Yes	Yes				
I think we can have a go.	Many	Yes	Yes	Yes			
Yep, I could use this	Many	Yes	Yes	Yes	Some		
Smoking	Many	Yes	Yes	Yes	Clear		
We were one of the early adopters ...	Many	Yes	Yes	Yes	Clear	Yes	
It's so mainstream now.	Many	Yes	Yes	Yes	Clear	Yes	Yes

Data source: <http://blog.gardeviance.org/2008/10/cloud-adoption-model.html>

