

All I Need is Scale!

Growing and Evolving a Large eCommerce Site

Franco Travostino
eBay Distinguished Architect

CISCO Cloud Computing Research Symposium

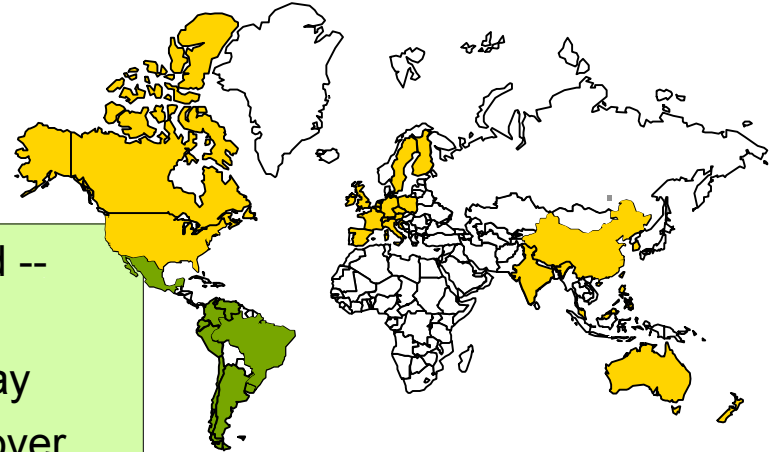
Cisco Systems
November 5th and 6th, 2008



Challenges at Internet Scale

- eBay manages ...
 - Over 276,000,000 registered users
 - Over 2 Billion photos
- eBay users trade \$2040 in goods every second -- \$60 billion per year
- eBay averages over 2 billion page views per day
- eBay has roughly 112 million items for sale in over 50,000 categories
- Over 7 million listings added per day
- eBay site stores over 2 Petabytes of data
- In a dynamic environment
 - 300+ features per quarter
 - We roll 100,000+ lines of code every two weeks
- In 39 countries, in 8 languages, 24x7x365

>48 Billion SQL executions/day!



Scaling dimensions

- Data
- Transactions
- SLAs
- Operations
- Deployment
- Productivity
- Feature velocity
- Adaptability

Five principles that served us well

1. Partition Everything

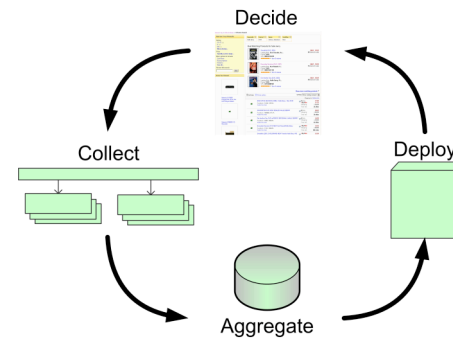
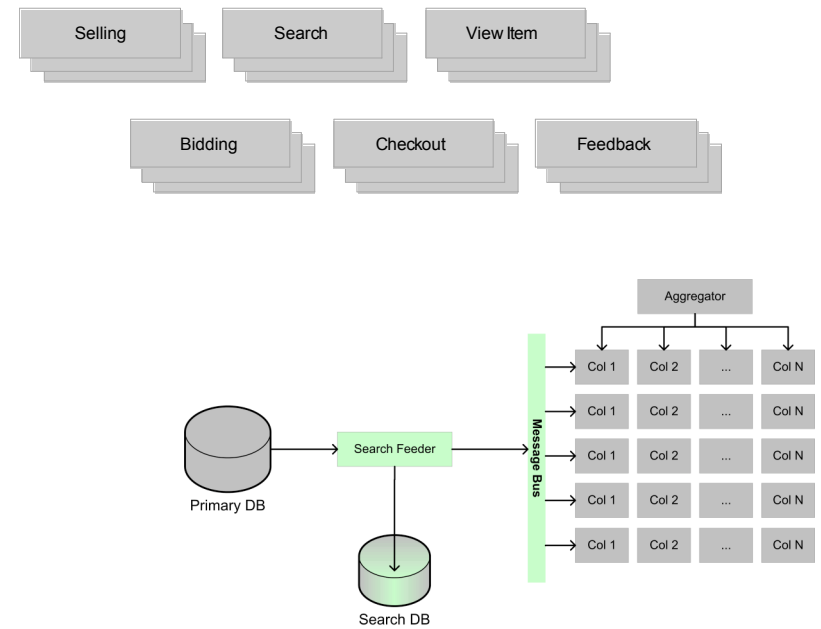
- Functional partitions
- Horizontal splits (“shards”)
- No session state in application tier

2. Asynchrony Everywhere

- Events (multi-topic, idempotency, ...)
- Multicast message dissemination

3. Automate Everything

- Adaptive configuration
- Increasingly, machine learning



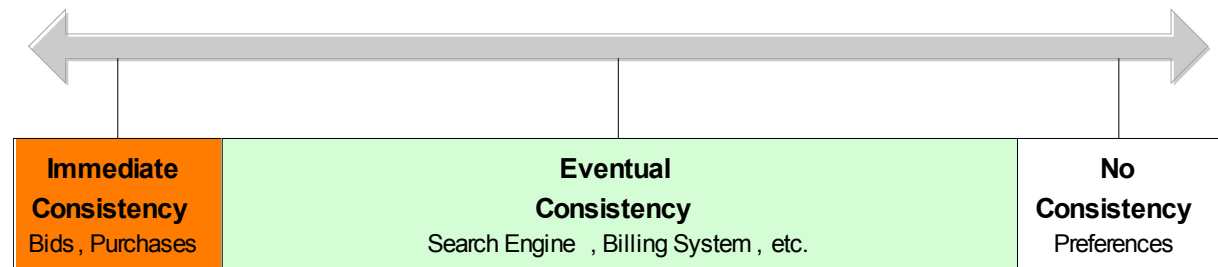
Five principles that served us well—cont'd

4. Remember Everything Fails

- Rollback - no change to the site unless it can be undone
- Extensive logs (*~2 TB of log messages a day*)
- Policies and mechanisms for graceful degradation

5. Embrace Inconsistency

- Consistency is a spectrum, not binary
- Use-case specific tradeoffs (CAP)
- Just say no to distributed transactions





'Odysseus and the Sirens'. Draper, 1909.

Clouds, a new chapter in the scale-out journey?

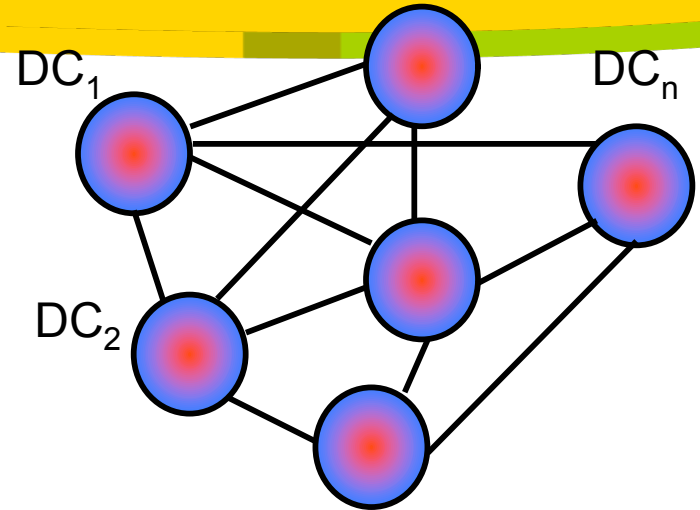
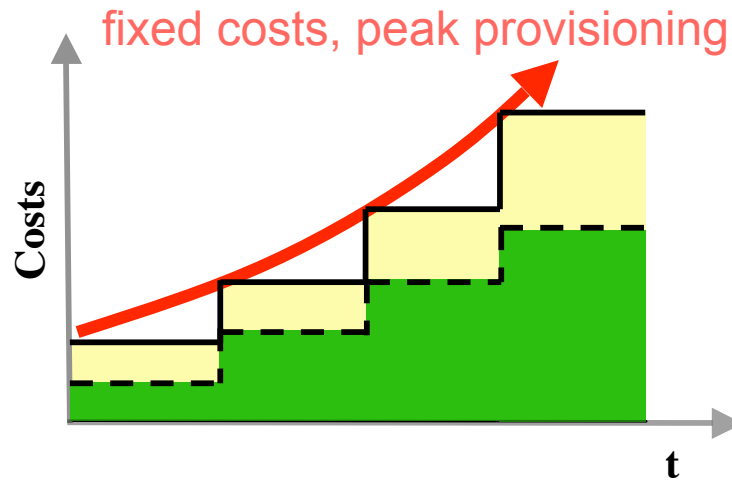
Had he had a choice in '95...



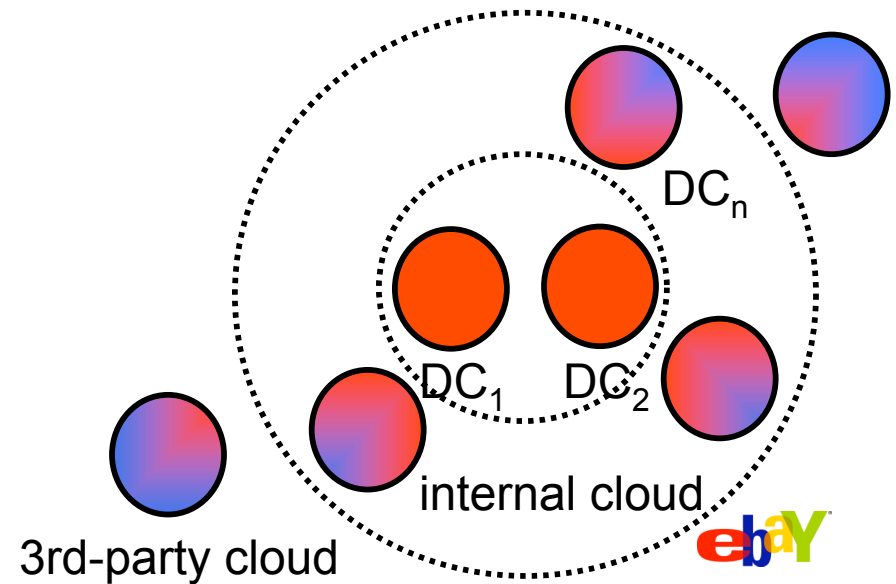
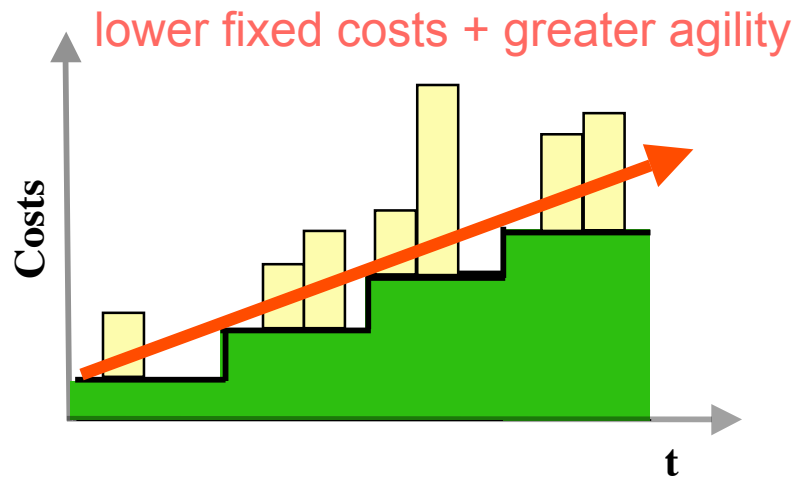
... would Pierre Omidyar have started eBay.com off a Cloud?

Quite likely!

Clouds for a F500 like eBay, today



Before
After



Economists knew it all along

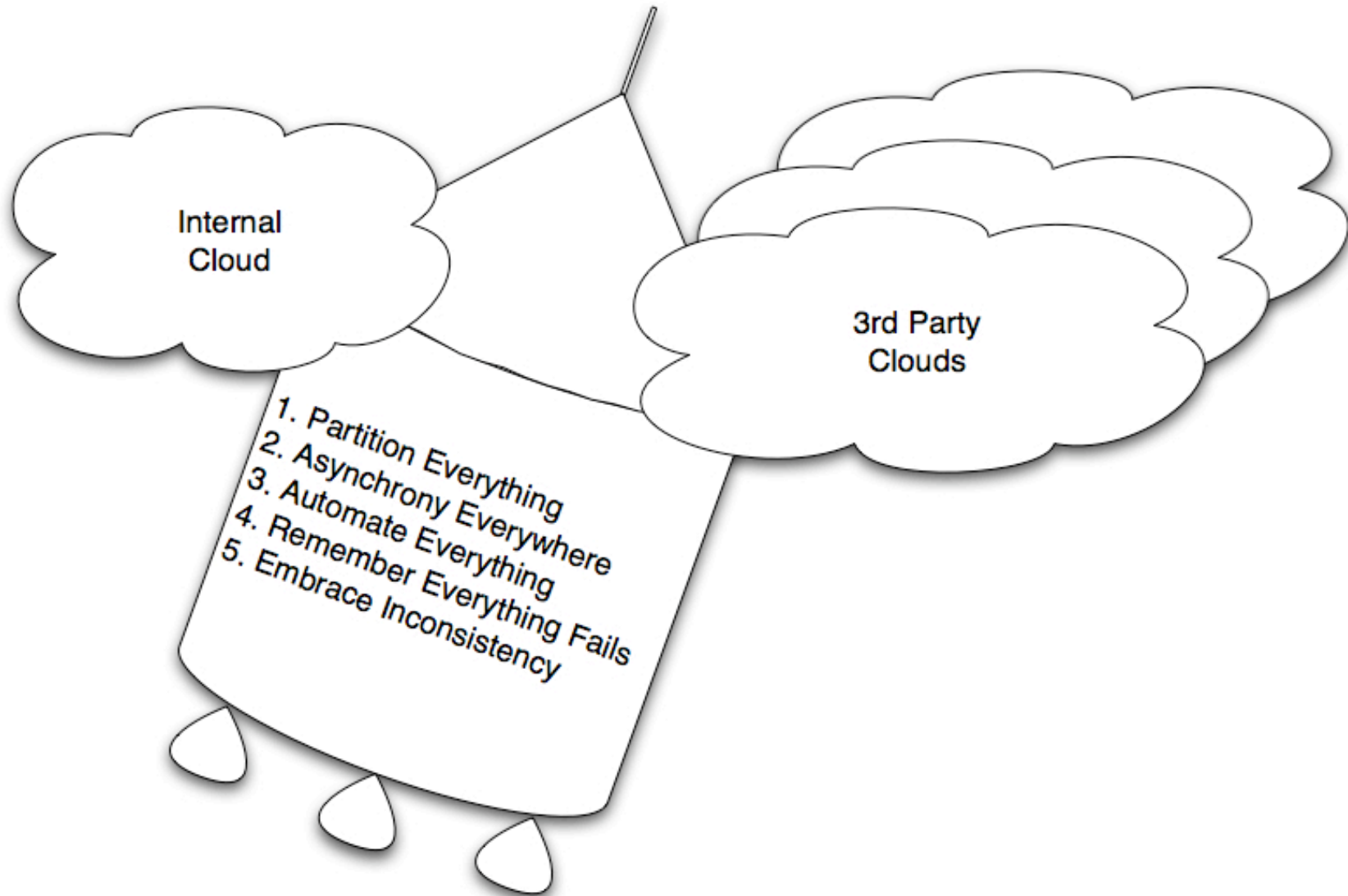


Ronald Coase
Nobel Laureat, 1991

"a firm will tend to expand until the costs of organizing an extra transaction within the firm become equal to the costs of carrying out the same transaction by means of an exchange on the the open market...."
-- *"The Nature of the Firm" (1937)*

- Same root as outsourcing and offshoring

Same five principles thrust us into the new journey

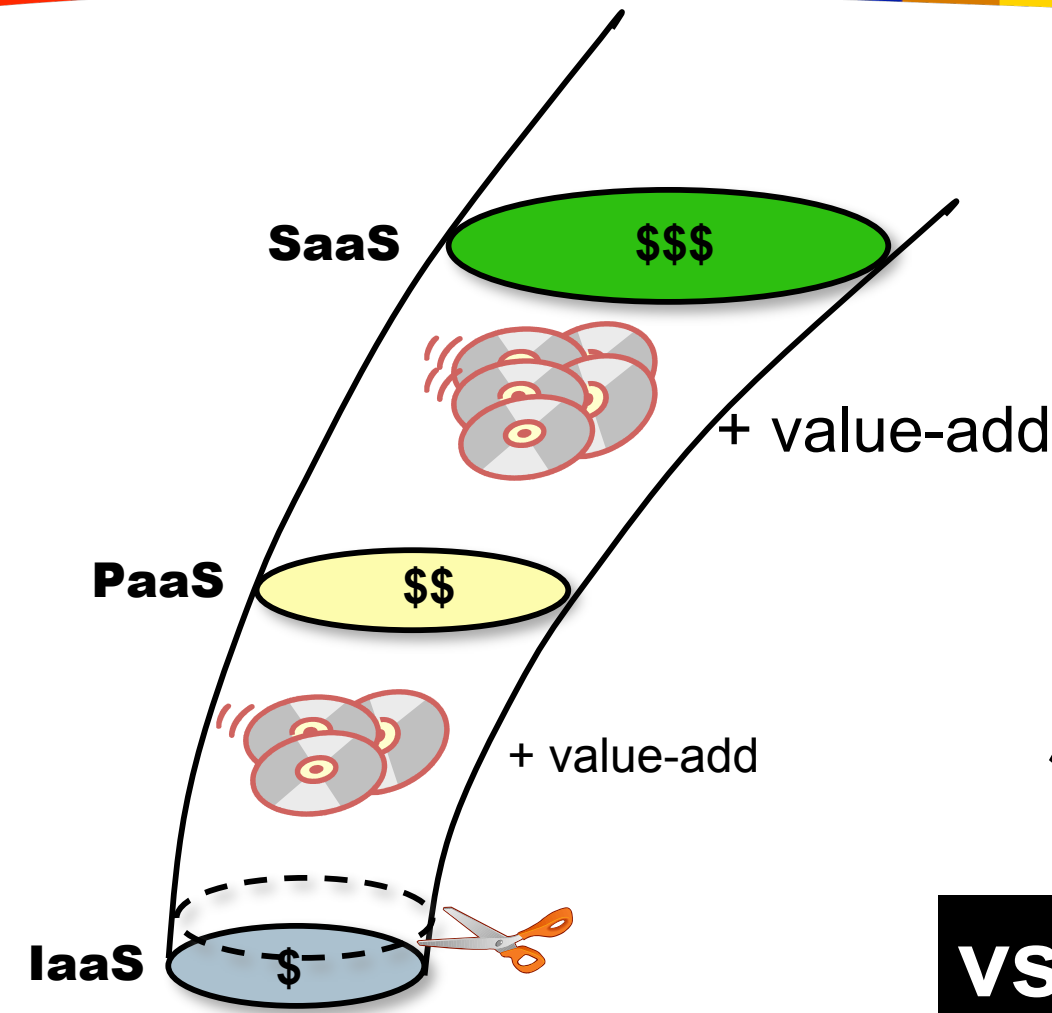




Faux PaaS [foh pah]: a slip or blunder in dealing with
Platform as a Service (or IaaS, SaaS)

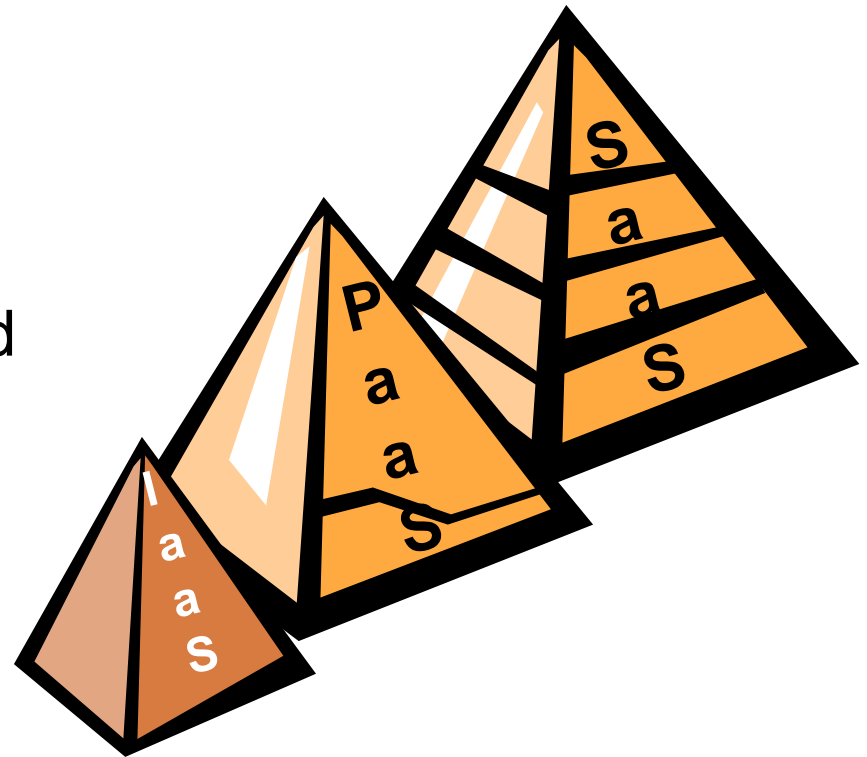


1st FauxPaaS: I'll move in or out of Clouds at whim



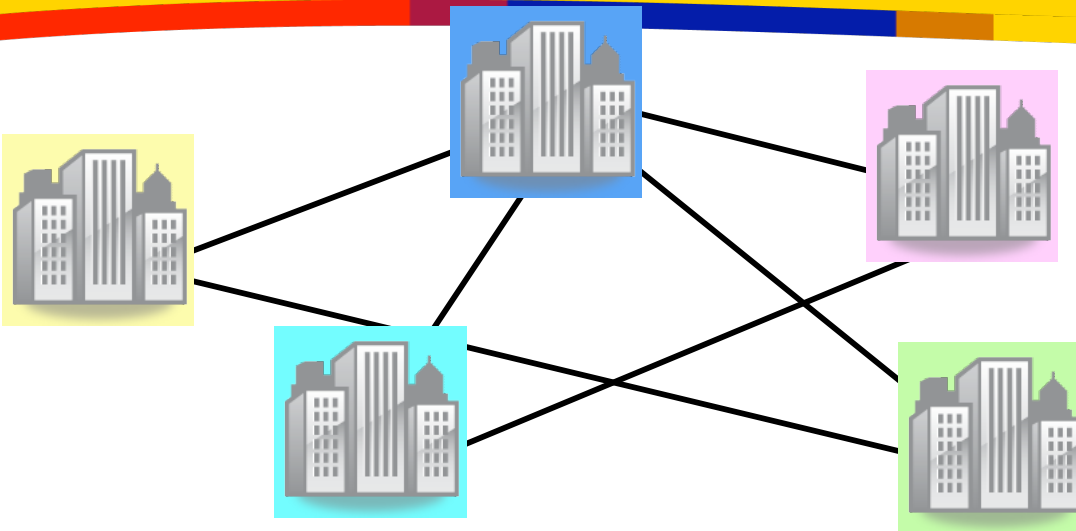
> Encapsulated, portable value-add

VS.



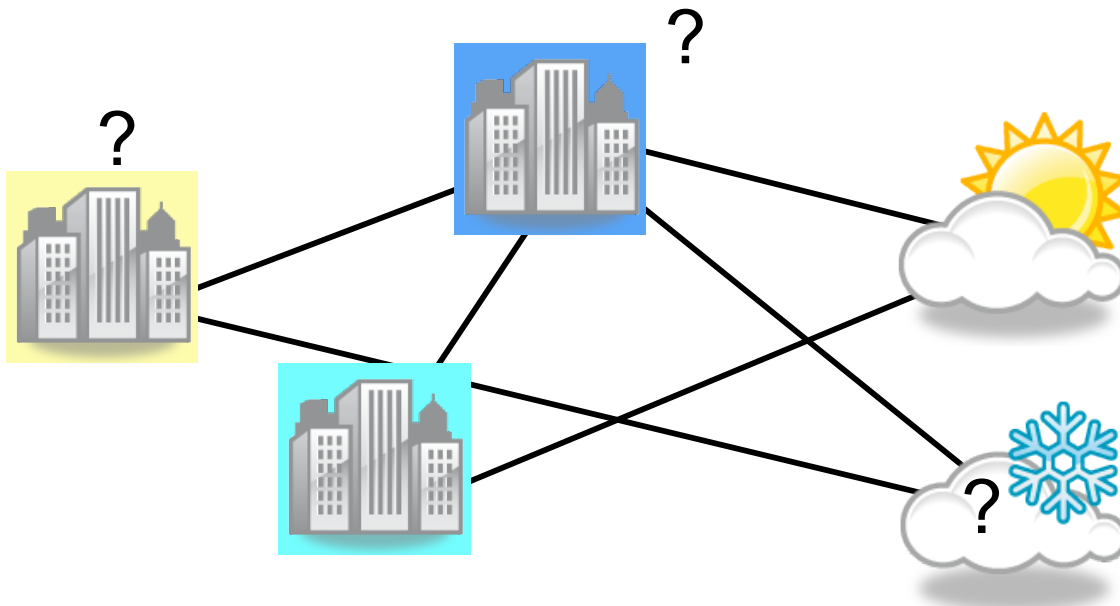
> Ad-hoc abstractions

2nd FauxPaaS: Cloud is B2B as usual



Traditional B2B setup:

- ✓ data federation
- ✓ distributed workflow



Generalize w/ Clouds:

- ✓ data federation
- ✓ distributed workflow
- ! no accountability
- ! soft SLAs

3rd FauxPaaS: Sitting idle waiting for The Data...

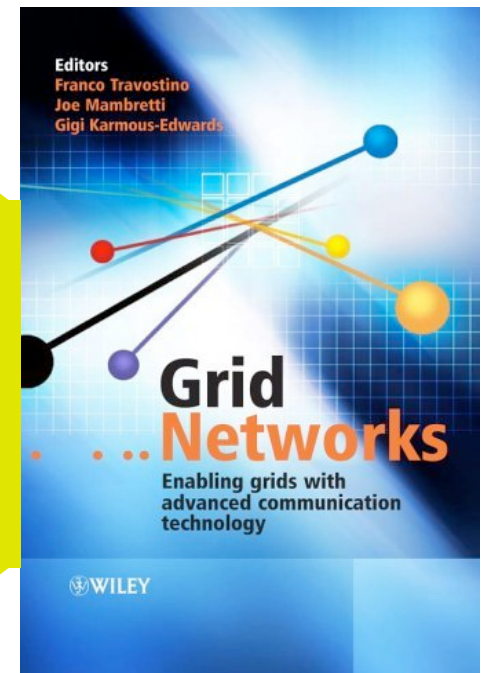
- Data access poses latency and/or throughput requirements to e2e network
- Network-aware Cloud economics
 - Jim Gray's Distributed Computing Economics (2003)
 - “Never underestimate the bandwidth of a station wagon full of tapes ...” (Andy T.)
- If network plays key role, why aren't Telcos stepping up to become Cloud dominant players?
- Hypothesis: slow start but they will, once biz matures

<plug>

* Some exciting network opportunities from Research

* s/Grid/Cloud/

</plug>



Research Challenges

- Novel languages/runtimes to harness many-core parallelism
- Model-driven architectures
 - Model reflects relationships, behaviors, constraints
 - Seamless off-cloud and on-cloud transition
- N-resource M-constraint scheduling
- Statistical multiplexing, service level management
- Cost attribution/recovery models
- Stronger trust models
- Energy-aware Computer Science
- Experience-driven design practices (e.g., planetlab)

Summary

- Five principles underlie eBay's scale-out journey to date
- Clouds can improve our cost models and rev up innovation velocity
 - Internal cloud
 - 3rd party clouds
- The same five principles and patterns continue to apply (more so than ever!)
- Will thread carefully and avoid *Faux PaaS...*
 - Lock-in
 - Soft SLAs, no accountability
 - Compute vs. data locality, is the network experience up to par?
- New research challenges hit on a variety of disciplines