Infrastructure Innovation Opportunities

Y Combinator 2013

James Hamilton, 2013/1/22
VP & Distinguished Engineer, Amazon Web Services
email: James@amazon.com
web: mvdirona.com/jrh/work
blog: perspectives.mvdirona.com
Agenda

- Costs Drive Startup Opportunity
- Networking
- Storage
- H/W Innovation
- Cloud Computing
  - Cloud Economics
  - 2\textsuperscript{nd} Tier Effects
Costs Drive Startup Opportunity

- **Assumptions:**
  - Facility: ~$88M for 8MW critical power
  - Servers: 46,000 @ $1.45k each
  - Commercial Power: ~$0.07/kWhr
  - Power Usage Effectiveness: 1.45

  ![Monthly Costs Pie Chart]

  - 57% Power
  - 18% Networking Equipment
  - 13% Power Distribution & Cooling
  - 4% Servers
  - 8% Other Infrastructure

- **Observations:**
  - 31% costs functionally related to power (trending up while server costs trending down)
  - Networking high at 8% of overall costs & 12% of total IT gear cost (many pay more)


2013/1/22
Sea Change in Networking

- Current networks over-subscribed
  - Forces workload placement restrictions
  - Goal: all points in datacenter equidistant
- Mainframe model goes commodity
  - Competition at each layer over vertical integ.
- Get networking onto Moore's Law path
  - ASIC port count growth at near constant cost
  - Competition: Broadcom, Marvell, Fulcrum,...

Key:
- CR = L3 Core Router
- AR = L3 Access Router
- S = L2 Switch
- LB = Load Balancer
- A = Rack of 20 servers with Top of Rack switch
Software-Defined Networks

Application
Application
Application
Application

Network Operating System

Simple Packet Forwarding Hardware

Simple Packet Forwarding Hardware

Simple Packet Forwarding Hardware

Simple Packet Forwarding Hardware

2013/1/22

http://perspectives.mvdirona.com
HDD Random BW vs Sequential BW

- Disk sequential BW growth slow
- Disk random access BW growth roughly 10% of sequential
- Storage chasm widening
  - BW a long term problem & IOPS growth very slow

Source: Dave Patterson with James Hamilton updates
Disk Becomes Tape

- Random disk latency increasingly impractical
- Sequential full 4TB read is over 11 hours
- Random full read 4TB disk:
  - 41.3 days @ 140 IOPS with 8kb page
  - Disk increasingly impractical for random workloads
- Cold storage biggest storage market
- Trending below tape price point
  - Tape only cost effective at very high scale
  - Disk wins at top and scales down better

Tape is Dead
Disk is Tape
Flash is Disk
RAM Locality is King

Jim Gray
Microsoft
December 2006
Flash Becomes Disk

• All random IOPS workloads to Flash
• Flash 4 to 6x more expensive by capacity
• Technique: log structured store
  – Compress
  – De-dupe
  – Sparse provision
• Approaches HDD capacity price point
Client Storage Migration

• Client device disk replaced by semiconductor caches
  – Much higher performance, Lower power dissipation, smaller form factor, greater shock resistance, scale down below HDD cost floor, greater humidity range, wider temp range, lower service costs, ...

• Clients storage drives cloud storage
  – Value added services, many data copies, shared access, indexed, classified, analyzed, monetized, reported, ...
  – Overall client storage continuing to expand rapidly but primarily off device in cloud
Practical to Innovate at Any Level

• Can’t afford a $4B to $8B fab
  – Don’t have to: TSMC, Global, Samsung, ...

• Can’t afford to write custom EDA tools
  – Don’t have to: Synopsys, Cadence, ...

• Can’t afford to do a custom processor design
  – Don’t have to: ARM license with custom IP blocks

• Can’t afford device manufacturing plant
  – Don’t have to: Foxconn, Quanta, Wistron, ....

• Can’t afford world-wide datacenters & all the servers in each
  – Don’t have to: AWS, Azure, GAE,...

• Can’t afford to build the entire s/w stack
  – Don’t have to: many active open source communities

• Smallest team can do custom devices & scalable service
The Cloud Changes Everything

- Scale economics up several orders of magnitude
- Infrastructure utilization key lever
- Data center Innovation & efficiency
- Custom, service-specific hardware
- Cloud: low-cost, very high-volume business
  - Not on enterprise uplift model
- Opportunities:
  - Infrastructure-free startups (and very large businesses)
  - 2nd tier effect
Perspective on Scaling

Each day, AWS adds enough server capacity to support all of Amazon’s global infrastructure in 2003 when it was a $5.2B annual revenue enterprise.
The Cloud Scales: Amazon S3 Growth

Peak Requests:
500,000+ per second

<table>
<thead>
<tr>
<th>Year</th>
<th>Total Number of S3 Objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q4 2006</td>
<td>2.9 Billion</td>
</tr>
<tr>
<td>Q4 2007</td>
<td>14 Billion</td>
</tr>
<tr>
<td>Q4 2008</td>
<td>40 Billion</td>
</tr>
<tr>
<td>Q4 2009</td>
<td>102 Billion</td>
</tr>
<tr>
<td>Q4 2010</td>
<td>262 Billion</td>
</tr>
<tr>
<td>Q4 2011</td>
<td>762 Billion</td>
</tr>
<tr>
<td>Q4 2012</td>
<td>&gt;1 Trillion</td>
</tr>
</tbody>
</table>

2013/1/22
http://perspectives.mvdirona.com
AWS Datacenters in 8 Regions

US GovCloud (US ITAR Region -- Oregon)
US West x 2 (N. California and Oregon)
US East (Northern Virginia)
Europe West (Dublin)
Asia Pacific Region (Singapore)
Asia Pacific Region (Tokyo)

LATAM (Sao Paola)

>10 datacenters In US East alone

8 AWS Regions and growing
21 AWS Edge Locations for CloudFront (CDN) & Route 53 (DNS)
Utilization & Economics

• Server utilization problem
  – 30% utilization VERY good & 10% to 20% common
    • Expensive & not good for environment
  – Solution: pool number of heterogeneous services
    • Non-correlated peaks & law of large numbers

• Pay as you go & pay as you grow model
  – Don’t block business
  – Don’t over buy
  – Transfers capital expense to variable expense
  – Apply capital for business investments rather than infrastructure

• Charge back models drive good application owner behavior
  – Cost encourages prioritization of work by application developers
  – High scale needed to make a market for low priority work
Data Center Efficiency

• Datacenter design efficiency
  – Average datacenter efficiency low with PUE over 2.0 (Source: EPA)
    • Many with PUE over 3.0
  – High-scale cloud services in 1.2 to 1.5 range
  – Lowers computing cost & better for environment

• Multiple datacenters
  – At scale multiple datacenters can be used
    • Close to customer
    • Cross datacenter data redundancy
    • Address international markets efficiently

• Avoid upfront datacenter cost with years to fully utilize
  – Scale supports pervasive automation investment
Hardware Scale Effects

- Custom service-optimized hardware
  - ODM sourced
- Purchasing power at volume
- Supply chain optimization
  - Shorter supply chain drives higher server utilization
    - Predicting next week easier than 4 to 6 months out
  - Less over buy & less capacity risk
- Networking transit costs strongly rewards volume
- Cloud services unblocks new business & growth
  - Remove dependence on precise capacity plan
Amazon Cycle of Innovation

• 15+ years of operational excellence
  – Managing secure, highly available, multi-datacenter infrastructure

• Experienced at low margin cycle of innovation:
  – Innovate
  – Listen to customers
  – Drive down costs & improve processes
  – Pass on value to customers

• 21 AWS price reductions so far
  – Expected to continue
2\textsuperscript{nd} Tier Provider Effect

• Amazon investments tend to be:
  – Early stage technology
  – Later stage companies with developed markets
  – Most AWS technology internally developed, but ...

• Internally developed AWS technology opens up startup sales & acquisition opportunities
  – Cloud market large with some companies not software focused
  – Leaders push innovation while 2\textsuperscript{nd} tier players buy or acquire
Questions?

• Perspectives Blog:
  – http://perspectives.mvdirona.com/

• Email:
  – James@amazon.com