Data Center Infrastructure Innovation

Leading Edge Forum San Francisco

James Hamilton, 2010.10.20 VP & Distinguished Engineer

e: James@amazon.com w: mvdirona.com/jrh/work b: perspectives.mvdirona.com



Agenda

- Quickening pace of DC infrastructure innovation
- Where does the money go?
- Power distribution infrastructure
- Mechanical systems
- · Sea change in net gear
- Server innovations
- Cloud Computing Economics







- Why utility computing makes sense economically









 Talk does not necessarily represent positions of current or past employers

 2010/10/20
 http://perspectives.mvdirona.com

Pace of Innovation

- Datacenter pace of innovation increasing
 - Driven by cloud service providers and very high scale internet applications like search
 - Cost of datacenter & H/W infrastructure dominates
 - Core business rather cost center $Google^{-}$
- High focus on infrastructure innovation
 - Driving down cost
 - Increasing aggregate reliability
 - Reducing resource consumption footprint

facebook

Microsof

Where Does the Money Go?

Assumptions:

- Facility: ~\$72M for 8MW critical load
- Servers: 46,000 @ \$1.45k each
- Commercial Power: ~\$0.07/kWhr
- Power Usage Efficiency: 1.45





3yr server & 10 yr infrastructure amortization

• Observations:

- 31% costs functionally related to power (trending up while server costs down)
- Networking high at 8% of costs & 19% of total server cost (many pay more)

 $From: \ http://perspectives.mvdirona.com/2010/09/18/OverallDataCenterCosts.aspx$

Power Distribution



http://perspectives.mvdirona.com

Mechanical Systems



Hot Aisle/Cold Aisle Containment





Intel



WriteLine

Intel

ASHRAE 2008 Recommended



ASHRAE Allowable



Air-Side Economization & Evaporative Cooling

- Operate with higher server inlet temps
- Limitations to high temp operation
 - Higher fan power trade-off
 - More semiconductor leakage current
 - Possible negative failure rate impact
- Avoid direct expansion cooling entirely
 - Air side economization
 - Higher data center temperatures
 - Evaporative cooling
- Requires Filtration
 - Particulate & chemical pollution





Sea Change in Net Gear

- Current networks over-subscribed
 - Forces workload placement restrictions
 - Goal: all points in datacenter equidistant
- Mainframe model goes commodity
 - Competition at each layer rather than vertical integration
- OpenFlow: open S/W platform
 - Distributed control plane to central control
 - E.g. VL2, Portland, and others





Central Logic Manufacture

Application Stack •Not supported •No programming tools •No 3rd party ecosystem

Net Equipment

Central Logic Manufacture •Standard design (x86) •Multiple source •AMD, Intel, Via, ... Finished Hardware Supply •Standard design •Multiple source •Dell, SGI, HP, IBM, ... System Software Supply •Linux (many distros/support) •Windows & other proprietary offerings

Application Stack •Public/published APIs •High quality prog tools •Rich 3rd party ecosystem

Commodity Server



http://perspectives.mvdirona.com

Server Innovation

- Shared Infrastructure Racks
 - Shared PSUs, NICs & fans
 - e.g. Dell Fortuna & Rackable CloudRack
- Next Level: Multi-server on board
 - Intel Atom: SeaMicro
 - ARM: SmoothStone
- Very Low-Cost, Low-Power Servers
 - ARM, Atom, client & embedded CPUs
 - Cold storage (reduce CPU \$ to GB)
 - Partitionable workloads: web servers, memcached
- Low utilization is still the elephant in room







Infrastructure at Scale

- Datacenter design efficiency
 - Average datacenter efficiency low with PUE over 2.0 (Source: EPA)
 - Many with PUE well over 3.0
 - High scale cloud services in the 1.2 to 1.4 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 massive upfront data cost & years to fully utilize
 - Scale supports pervasive automation investment

Utilization & Economics

- Server utilization problem
 - 30% utilization VERY good &10% to 20% common
 - Expensive & not good for environment
 - Solution: pool number of heterogeneous services
 - Single reserve capacity pool far more efficient
 - Non-correlated peaks & law of large numbers
- Pay as you go & pay as you grow model
 - Don't block the 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





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
- AWS price reductions frequent & expected to continue

AWS Approach

- Broad set of services:
 - Infrastructure Services
 - SimpleDB
 - Simple Storage Service
 - CloudFront
 - Simple Queue Service
 - Elastic MapReduce
 - Relational Database Service
 - Elastic Block Store
 - Premium Support
 - Virtual Private Cloud

- Payments & Billing
 - Flexible Payment Services
 - DevPay
- On Demand Workforce
 - Mechanical Turk
- Alexa Web Services
 - Web Information Service
 - Top Sites
- Merchant Services
 - Fulfillment Web Service
- "Open the hood" approach
 - Simple, layerable building block services
 - Component services are substitutable





H/W Cost & Efficiency Optimization

- Service optimized hardware
 - Custom cloud-scale design teams:
 - Contract manufacturers, Dell DCS, Rackable, ZT Systems, HP, ...
- Purchasing power in volume
- Supply chain optimization
 - Shorter chain drives much higher server utilization
 - Predicting next week easier than 4 to 6 months out
 - Less overbuy & less capacity risk
- Networking transit costs strongly rewards volume
- Cloud services unblocks new business & growth
 - Remove dependence on precise capacity plan





AWS Pace of Innovation

» Amazon CloudFront Default Root » Amazon RDS: Northern California Region, Multi-AZ Objects Deployments, AWS Management Console support » Amazon RDS Reserved Instances » Amazon CloudFront: Access logs for streaming » Amazon CloudFront Invalidation » Amazon S3 Reduced Redundancy Storage » Amazon EC2 with Windows Server 2008, Spot Instances. Boot from Amazon EBS » Amazon CloudFront: » Amazon CloudFront Streaming HTTPS support, lower » Amazon RDS Read » Amazon VPC enters Unlimited Beta Replicas; lowers request pricing, NYC » AWS Region in Northern California » AWS SDK for Java edge location prices » International Support for AWS Import/Export » Windows server 'Bring Your Own » AWS Import/Export exits » Amazon EC2 running License' pilot program beta; web service support SUSE Linux » Amazon CloudFront: Singapore » AWS Console » AWS Management edge location, private content for » Amazon RDS supports Amazon Console for Amazon S3 » High-Memory Instances streaming SNS » Amazon CloudWatch » Lower EC2 Pricing » Amazon ELB support monitoring for Amazon for HTTPS **EBS** volumes 2009 2010 Sep Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Oct » Amazon SQS Free Tier » Amazon EC2 Reserved Instances » Amazon S3 Bucket Policies with Windows, Extra Large High » EBS Shared » Amazon VPC IP Address Assignment Memory Instances **Snapshots** » Amazon EC2 Cluster Compute Instances » Amazon S3 Versioning Feature » SimpleDB in EU » Amazon S3 Enhanced Support for RRS » Consolidated Billing for AWS Region » Lower Amazon EC2 Pricing » Lower pricing for Outbound Data » Monitoring, Auto » AWS IAM Preview Beta Transfer Scaling & Elastic » AWS Console Support for Amazon VPC » Combined AWS Data Transfer Load Balancing in » Amazon EC2 Micro Instances Pricing EU » Amazon Linux AMI » Amazon SNS » Amazon EC2 Tagging, Filtering, Import Key » Amazon CloudFront » Amazon Elastic MapReduce: custom Pair, Idempotency Private Content cluster configuration option » Oracle certifies enterprise software on Amazon » SAS70 Type II Audit » Amazon RDS: EU Region launch EC2 » AWS SDK for .NET » AWS Asia Pacific (Singapore) » AWS SDK on PHP

Region

More Information

- These Slides:
 - I'll post the slides to <u>http://mvdirona.com/jrh/work</u> later this week
- Power and Total Power Usage Effectiveness
 - http://perspectives.mvdirona.com/2009/06/15/PUEAndTotalPowerUsageEfficiencyTPUE.aspx
- Berkeley Above the Clouds Paper
 - <u>http://perspectives.mvdirona.com/2009/02/13/BerkeleyAboveTheClouds.aspx</u>
- Degraded Operations Mode
 - <u>http://perspectives.mvdirona.com/2008/08/31/DegradedOperationsMode.aspx</u>
- Cost of Power
 - <u>http://perspectives.mvdirona.com/2008/11/28/CostOfPowerInLargeScaleDataCenters.aspx</u>
 - <u>http://perspectives.mvdirona.com/2008/12/06/AnnualFullyBurdenedCostOfPower.aspx</u>
- Power Optimization
 - <u>http://labs.google.com/papers/power_provisioning.pdf</u>
- Cooperative, Expendable, Microslice Servers
 - <u>http://perspectives.mvdirona.com/2009/01/15/TheCaseForLowCostLowPowerServers.aspx</u>
- Power Proportionality
 - <u>http://www.barroso.org/publications/ieee_computer07.pdf</u>
- Resource Consumption Shaping:
 - <u>http://perspectives.mvdirona.com/2008/12/17/ResourceConsumptionShaping.aspx</u>
- Email & Blog
 - <u>James@amazon.com</u> & <u>http://perspectives.mvdirona.com</u>