The Green Cloud: How Cloud Computing Can Reduce Datacenter Power Consumption

Anne M. Holler

Senior Staff Engineer, Resource Management Team, VMware

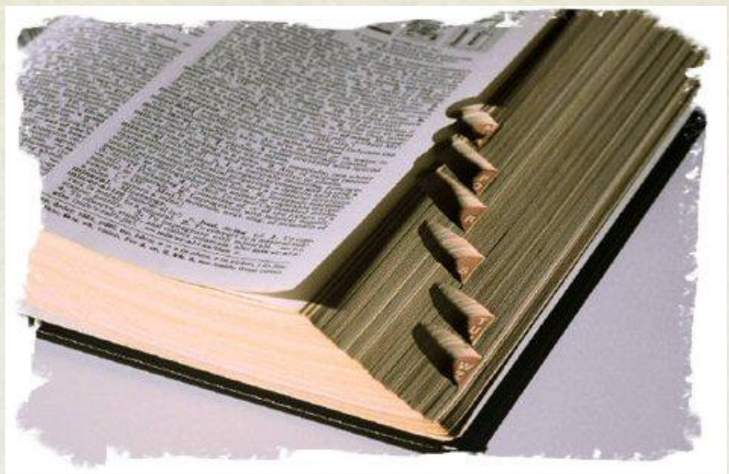
Foreword

- Datacenter (DC) energy consumption is significant
  - In 2006, it was 1.5% of all US energy consumption, projected at that point to double by 2011 [USEPA07b]
  - DC per-server power consumption cost over lifetime now is typically higher than server purchase cost [BELA07]

Foreword

- Talk envisions a future point in cloud computing
  - Describes idealized functionality not limited to that available in currently shipping products
  - Assumes solutions to cloud computing adoption challenges (e.g., security)

Cloud Computing [NIST09]

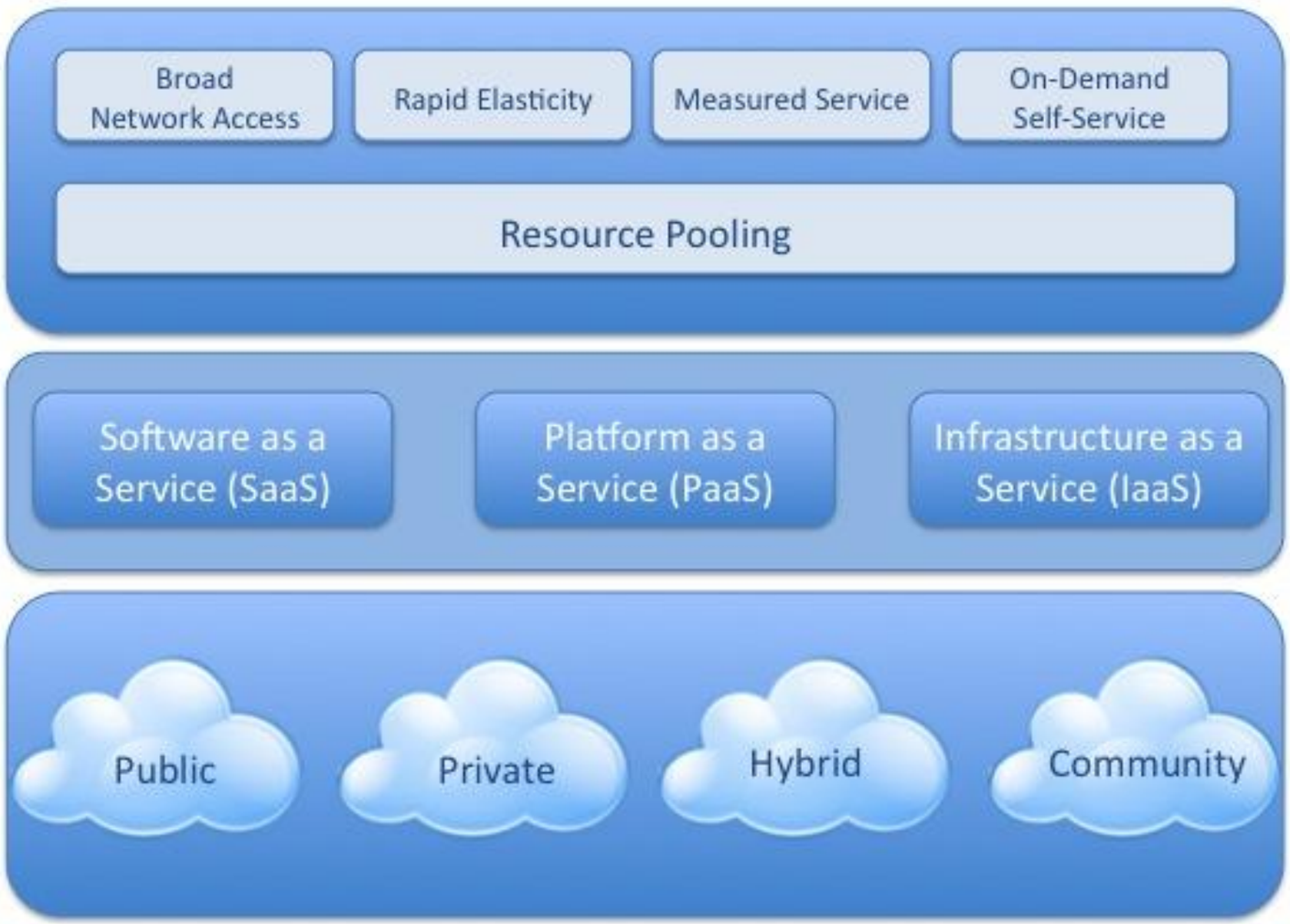


- Definition: “Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.”

Cloud Computing [NIST09]

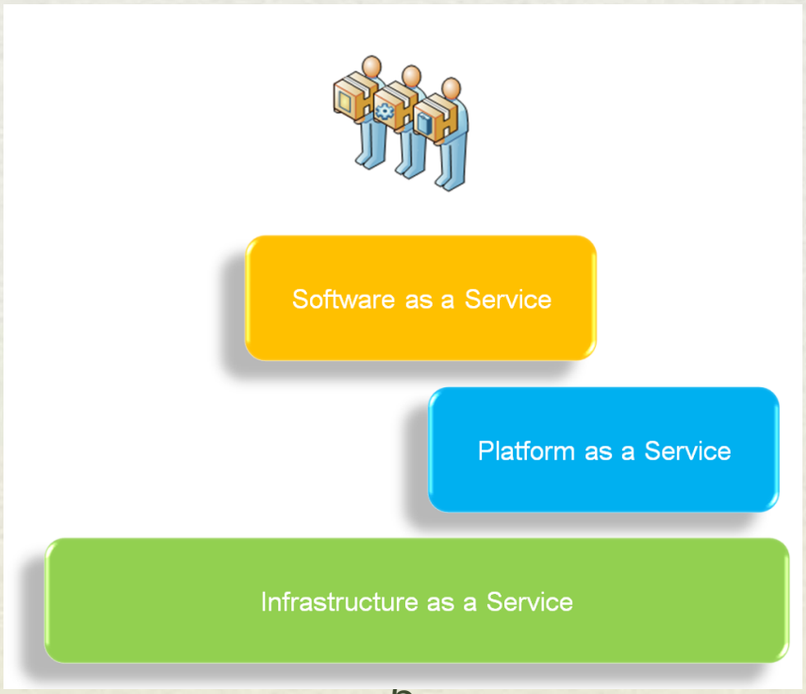
Visual Model Of NIST Working Definition Of Cloud Computing

http://www.csrc.nist.gov/groups/SNS/cloud-computing/index.html



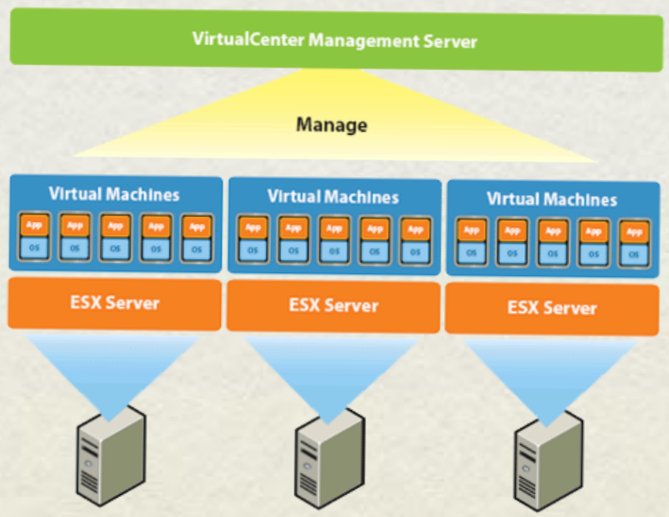
Focus on Cloud IaaS

- Talk focusses on Infrastructure as a Service (IaaS) model
  - Lowest service level on which other levels may be built
- IaaS is often based on DC virtualization



DC Virtualization for IaaS

- IaaS needs a standard framework for running diverse workloads; Virtual machine (VM) application & OS encapsulation provides this
- IaaS needs method for resource pooling for flexible multi-tenancy; VM Consolidation on fewer physical hosts than native enables this
- IaaS needs to supply measured service and rapid elasticity; Virtualization Resource Management (RM) supports this

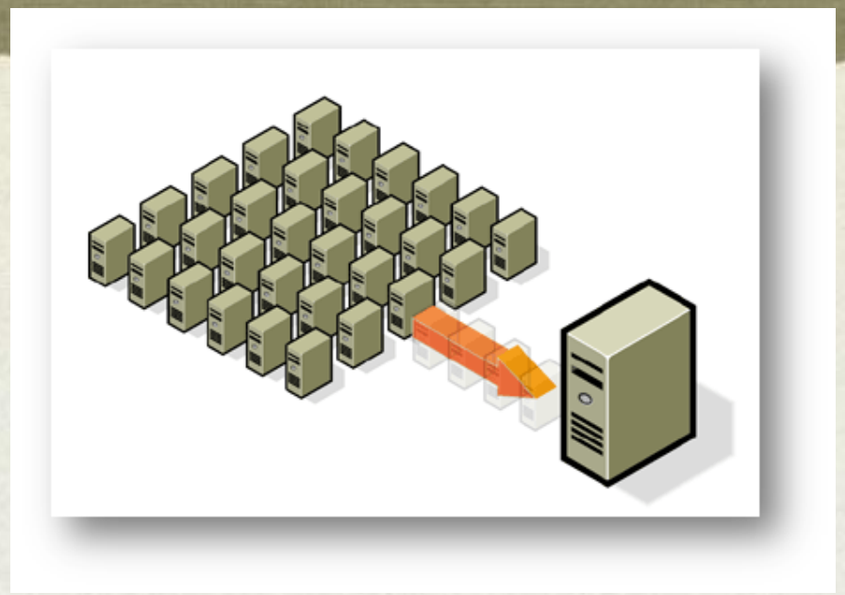


VMware Virtual Center provides a central point of control for managing, monitoring, provisioning and migrating virtual machines.

Outline

- How cloud computing can reduce DC power consumption
  - How DC virtualization (basis for IaaS) can save power
  - How cloud model can increase that power savings
- Then brief counter-arguments
  - How cloud computing can increase DC power consumption

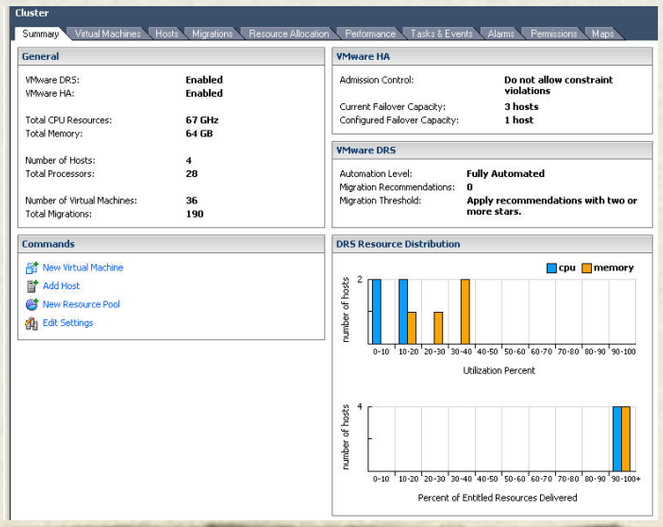
VM Consolidation Saves Power



- Volume physical hosts often 5-15% utilized [USEPA07a], while consuming 60-90% of their peak power [BODIK06]
- Power savings varies; est. 80% in [VMWARE08]; YMMV
- Utilities rebates for virtualization reducing power & hosts, e.g., PG&E pays $200/host removed, est. $300/host direct savings/year (~2x if cooling included) [PGE09]

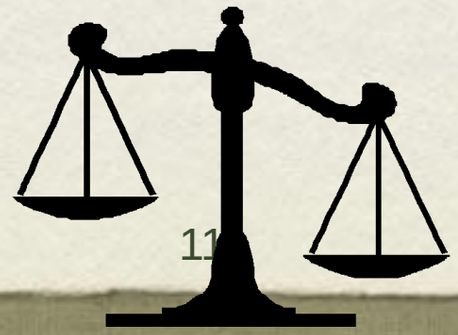
Virtualization RM for IaaS

- Consolidation depends on effective DC Virtualization RM
  - Uses resources efficiently to achieve high overall throughput
  - Allocates VM resources as per specified quality of service (QoS)
- Avoids undesirable cross-workload performance impact, e.g., Noisy neighbor issues in the cloud [NOISY10]



Virtualization RM for IaaS

- Elements of effective DC virtualization RM
  - Dynamic work-conserving allocation, not static partitioning
  - Rich resource control set, e.g., reservations, shares, limits, across all resources: CPU, memory, storage, network
- Hypervisor w/RM and isolation [e.g., VMware ESX]
- RM supported across cluster of hosts [e.g., VMware DRS]Using live migration btw hosts [e.g., VMware VMotion]

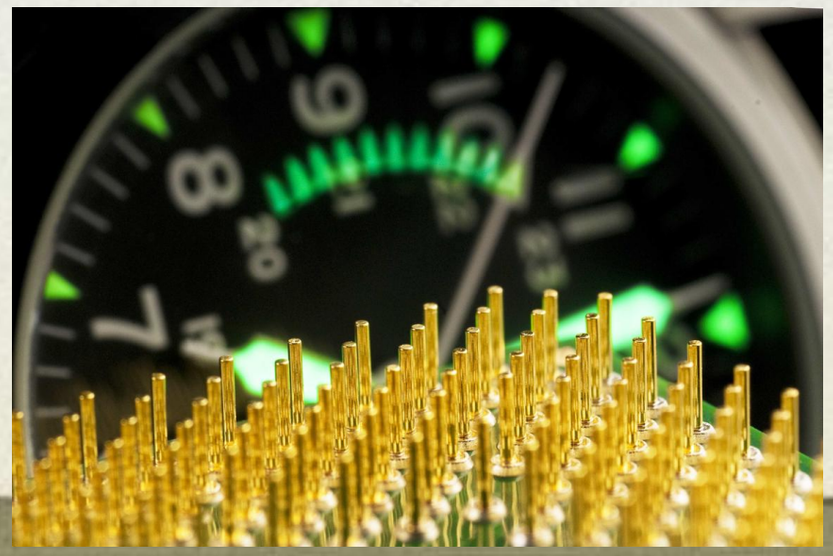


Cloud Adds Power Savings

- For private dedicated DCs already using DC virtualization, can the cloud model further increase power savings?
- Yes, because cloud computing model tends to foster:
  - Increased opportunities for VM consolidation
  - Increased incentive to reduce operating expenses (OpEx)

Cloud VM Consolidation

- Cloud provides aggregation point for workloads that would otherwise be run in separate DCs
- More pooling of workloads means more opportunity for:
  - Statistical multiplexing of demand via VM consolidation
  - Better exploitation of wider (cores) hosts, which tend to be more power-efficient than narrower ones [BARR05]



Dedicated DC Incentives

- Administrators of a private dedicated (non-cloud) DC often:
  - are not held accountable for power consumption OpEx
  - are held accountable for high resource availability
- Hence, low incentive to reduce power use and significant incentive to over-provision for peak (or even freak) usage

Dedicated DC Incentives

- Users of a private dedicated (non-cloud) DC often:
  - feel entitled to use(waste) resources based on their capital expenditures (CapEx); “all X hosts are belong to us” :-)
  - are not “billed” for resources their VMs consume
- Hence, low incentive to reduce VM resource consumption, in steady state or at peak

Cloud Provider Incentives

- Cloud provider wants low OpEx, for efficient resource pooling:
  - High utilization of powered-on computer resources
  - Low cost for computer resources kept as spare for peak demand
- Cloud provider wants low CapEx, limit DC maximum capacity
- Lead cloud provider to more aggressive methods to reduce OpEx and CapEx than those typical of dedicated DC administrator

Provider Reducing OpEx

- Power off hosts when demand low and back up when demand increases [e.g., VMware DRS w/DPM enabled]
  - Choose less efficient hosts to turn off, power-manage those on
  - Can save significant power, [DPM08] shows 55%; YMMV
  - For cyclical steep demand increase ("8am"): use demand prediction & proactive RM; for unpredictable increase ("slashdot"): use trend to spark proactive RM, reduce SLA

Provider Reducing OpEx

- Discount off-peak usage to promote use of more efficient hosts
  - Also reduces CapEx needed to handle maximum peak
- For multisite cloud providers, “Follow the moon”, i.e., move workloads to site providing cheapest power (storage VMotion)

Cloud User Incentives

- Cloud user wants low OpEx, basis to trade CapEx for OpEx:
  - Resource pooling removes CapEx resource entitlement
  - Measured service model highlights OpEx costs
  - Rapid elasticity avoids OpEx for peak usage until needed
- Leads cloud user to more aggressive methods to reduce OpEx than those typical of dedicated DC user

User Reducing OpEx

- Characterize computing resources needed for application workload to set appropriate cloud computing QoS SLA
  - Incorporate resource usage information in evaluating workload ROI & cloud provider choice
  - Determine response to workload demand changes, e.g., deployment specification in VMware vApp [SPRING09]
- Run non-time-critical workloads at off-peak times

User Reducing OpEx

- Reduce resource usage of application workload
  - Right-size VM [e.g.: match number vCPUs to workload parallelism, choose appropriate memory size]
  - Strip VM of unnecessary processes and services
  - Reduce usage when workload light/idle, e.g.: use tick-less OS, enable guest ACPI S1 sleep

Summary

- Cloud model can reduce DC power consumption by:
  - Workload consolidation via DC virtualization
    - Increased in cloud via statistical multiplexing
    - Incentives leading to aggressively lowering OpEx
- Also improves sustainability by reducing host count

Counter-Arguments

- How cloud model can increase DC power consumption
  - By facilitating execution of large workloads w/o CapEx, actual number of large DC workloads may grow
  - By increasing DC workloads' network accessibility, remote devices may proliferate, increasing aggregate power use
  - Cloud customers may expect higher SLAs from Cloud providers than they expect from their own dedicated DCs

Backup: References

- [BELA07] Belady, Christian. “In the Data Center, Power and Cooling Costs more than the IT Equipment it Supports”. Electronics Cooling, Vol. 23, No. 1, February 2007.

- [BODIK06] Bodik, P., M. Armbrust, K. Canini, A. Fox, M. Jordan, and D. Patterson. 2006. “A Case for Adaptive Datacenters to Conserve Energy and Improve Reliability”. Berkeley RAD Systems Laboratory. http://radlab.cs.berkeley.edu/

- [DPM08] http://www.youtube.com/watch?v=7CbRS0GGuNc

- [BARR05] Barroso, Luiz. “The Price of Performance”, ACM Queue, October 2005.

Backup: References

- [NIST09] http://csrc.nist.gov/groups/SNS/cloud-computing/

- [NOISY10] http://alan.blog-city.com/has_amazon_ec2_become_over_subscribed.htm

- [PGE09] Server Virtualization & Consolidation Fact Sheet http://www.pge.com/includes/docs/pdfs/mybusiness/energysavingsrebates/ incentivesbyindustry/hightech/C-4166.pdf- 188.1KB

Backup: References

- [SPRING09] http://blog.springsource.com/2009/08/13/virtualization-enterprise-java/

- [USEPA07a] US EPA. 2007. Working Group Notes from the EPA Technical Workshop on Energy Efficient Servers and Datacenters. Santa Clara, CA: U.S. Environmental Protection Agency. February 16.

- [USEPA07b] US EPA. 2007.Report to Congress on Server and Data Center Energy Efficiency Public Law 109-431. Report to Congress. U.S. Environmental Protection Agency. August 2.

- [VMWARE08] http://www.vmware.com/files/pdf/WhitePaper_ReducePowerConsumption.pdf