Cloud Computing & HP

Vanish Talwar
HP Labs
Cloud Computing – A Big Buzz Today

The Internet

The Web

The Cloud

virtualized services

information & e-commerce

connectivity

reach
But, what is Cloud Computing?

The definition is suitably vague. Some attempts:

**The 451 Group:** “The cloud is IT as a Service, delivered by IT resources that are independent of location”

**Gartner:** “Cloud computing is a style of computing where massively scalable IT-related capabilities are provided ‘as a service’ across the Internet to multiple external customers”

**Forrester:** “A pool of abstracted, highly scalable, and managed infrastructure capable of hosting end-customer applications and billed by consumption”
Cloud Embodies a Confluence of Technologies and Concepts

• Grid computing, utility computing, virtualization, SOA

• Direct comparisons with these technologies are ‘apples and oranges’ (or fruitless 😊) ...

• because Cloud computing is a conceptual service model, where:
  - Services are delivered remotely from a logical resource
  - Are paid for based on how much service is consumed
  - Are genuinely on-demand

• Cloud computing is a real trend driven by
  - The ubiquity of internet connectivity
  - Low-cost commodity hardware and open source software
  - Figuring out a bunch of technical stuff
Why we care about the Cloud

• The Cloud will transform the tech world
• Surpassing existing approaches that users and markets
  − can’t afford
  − can’t master
  − can’t access
• Changing the world for Users
• Changing the world for enterprises
What does the cloud add for Users?

- Cloud services can capture state and provide analytics to identify intentions
- Offer assistance
- Providing continuity and consistency across devices and services
- Requirements: persistence, scale, low cost
What does the cloud add for Enterprises?

- Cloud services enable handling millions of users
- Treating each as an individual
- Integrating with value network/ecosystem
- Allowing higher touch, increased relevance, better outcomes
- Resulting in growth and differentiation
- Requirements: persistence, scale, low cost
Designing the Cloud

- Business Models
- Social Models
- Technology
The Cloud Defined

Cloud Infrastructure

- Self Service
- Linear scaling, scale out, Flexing
- Variable payment
- Data analytics
The Cloud Defined

Cloud Services
- Low touch (credit card, not direct sales)
- Business, Developer, Infrastructure, Consumer Services
- Accessed over Internet, typically via web browser

Cloud Infrastructure
- Self Service
- Linear scaling, scale out, Flexing
- Variable payment
- Data analytics
The Cloud Defined

Cloud Services
- Low touch (credit card, not direct sales)
- Business, Developer, Infrastructure, Consumer Services
- Accessed over Internet, typically via web browser

Cloud Platforms
- Horizontal (Java, .NET, Open Source (Python etc))
- Vertical/Domain (SalesForce.com, Internet Retail, K-12 etc)
- Anchor ecosystems, value to users and service providers

Cloud Computing
- Self Service
- Linear scaling, scale out, Flexing
- Variable payment
- Data analytics

11 21 August 2008
Traditional Enterprise Approach = Distributed Design Responsibilities
Cloud Service Delivery = Integrated Design Responsibility

- Compute resources
- Open Source OS/middleware
- Application software
- Integrated System
Cloud Platform Design Target

**Scale**
- 1M-100M+ users

**Multi-tenancy**
- Support multiple customers on single instance

**Rich analytics**
- Comprehensive and unified generation of business and operations data
If an enterprise wants to lower IT costs, is Cloud the answer?

- If the app doesn’t benefit from scale and flexing, better answers are:
  - BTO
  - Automation and virtualization
  - HP Outsourcing
  - HP Flexible Computing Services
  - HP Adaptive Infrastructure
  - ...

- How about an internal cloud?
  - A fancy new name for High Performance Computing?
  - If you run it it’s not the Cloud
Is all hosting the Cloud?

• NO

• Depends on
  – Sales motion
  – Billing granularity
  – Time commitment
  – Instantiation/tear down time
  – Designed to scale
  – Need to flex
SaaS == Cloud?

- Not necessarily
- Depends on app architecture
  - Multi-tenancy?
  - Scale?
  - Analytics?
- Plus the other hosting questions
IT as a Service, Delivered by the Cloud
Cloud Service Varieties

Cloud End-User Services
Example: SalesForce.com CRM

Cloud Platform Services
Example: Google App Engine

Cloud Infrastructure Services
Example: Amazon EC2, S3

Physical Infrastructure
(not a cloud service)
Another Viewpoint: Cloud = Connecting

- People to people
- Business to businesses
- Experiences
- Businesses to customers
- Data to context and location

Cloud Ecosystem
Connecting people to people

Synchronization, Replication—Confusion!
Connecting people to people: Cloud
Connecting businesses: Traditional Approach

Your Business

Integration 1
Integration 2
Integration 3
Integration 4

Partner 1
Partner 2
Partner 3
Partner 4
Connecting businesses: Cloud

Your biz

integration

Partner 1

Partner 2

Partner 3

Partner 4

21 August 2008
Connecting Experiences
“We have 20 mins before landing, the AV system will now be switched off”
Imagine

John’s context in the Cloud
(ubiquitous, massive, unstructured and his)

event<.....
“jean de florette”,
frame(5658), ...
>

www.manley.john.cx
Later that day ....

... at the Downtown Los Angeles Marriott

The entertainment system offers to continue the movie!
A Loosely-coupled “App” . . .

get(Paused_Media, 3434W87T293G479hsjd4fh888kd47)

www.manley.john.cx

<Jean de Florette, VS-Edit, frame(5658):

www.imdb.com/streaming_service

enable “get” event
Print publishing:
An example of a cloud ecosystem
Print Publishing: Ecosystem
Print Publishing: Traditional I.T. Approach
Print Publishing: **Cloud Approach**
IT Challenge: Service Management in a Hybrid World
What we’ll do today

• Part I: Cloud Computing
  – Context and definitions
  – Cloud versus Traditional IT: What’s going on?
  – Architectural Challenges
  – Cloud = Connecting

• Part II: Research Opportunities, HP Labs Projects
  – Infrastructure for the cloud, disaggregated dematerialized datacenters
  – Sustainable IT Ecosystem
  – Cell as a Service
  – Service Lifecycle Management (SLiM)
  – Mercado: Business Operating Environment in Global Services Ecosystem
Cloud Research opportunities

**Cloud Services**
- Fitting technology into non-technology user experiences
- Service-centric Model-view-controller decomposition
- Delegation

**Cloud Platforms**
- Relational data models
- Content management
- Analytics
- Multi-tenancy

**Cloud Infrastructure**
- Software languages, runtimes, design patterns for concurrency
- Scale-out software design patterns for linear scaling
HP Labs High Impact Research Themes

Addressing the next technology challenges and opportunities

- Information explosion
- Cloud services
- Sustainability
- Content transformation
- Intelligent infrastructure
HP Labs: *’ing the Cloud

extending the Cloud
- Mercado
  
managing the Cloud
- SLiM (Service Lifecycle Management)
- Olympus (Scalable Monitoring & Platforms Mgmt.)

seeding the Cloud
  
forming the Cloud
- Cells-as-a-Service
- Infrastructure for the cloud
- Sustainable IT Ecosystem

massive, global, federated infrastructure
Infrastructure for the cloud
disaggregated dematerialized datacenters

Part of broader exascale datacenter project
- Disaggregated dematerialized datacenters
- Coordinated cross-layer management

Microblades & Megaservers for perf/$ improvements
- New benchmarks, metrics, cost/power/performance models
- Non-server CPUs, new packaging, ensemble memory sharing & flash

Coordination for more effective management
- Cross-layer across platforms and virtualization domains
- Better energy efficiency and meeting of customer SLA requirements
Scalable monitoring and platforms mgmt.

**Target System**
- Several millions of compute nodes
- Cutting across HW-SW boundaries

**Research Focus**
- Distributed overlays – discovery, directory services for mgmt. needs
- Adaptive and prediction-based monitoring
- Scalable power management

[Diagram of network topology with query response latencies and network latency metrics.]
Sustainable IT Ecosystem

Goal:
Creating technologies, IT infrastructure and business models for the lower-carbon economy – that save money and leave a lighter footprint.

Research focus:
Reinventing the ecosystem of billions of devices, millions of systems, and thousands of print factories and data centers

Result:
Reduction in consumption of materials
Less use of available energy
Fewer greenhouse gas emissions
Lower total cost of ownership
Growth of IT into new ecosystems
More users having access to IT services
Higher quality of IT services
Cells as a Service

• Create secure, dynamic virtual infrastructures (Cells) to host cloud-based services

• Research contributions
  – Cell isolation for security and performance
  – Cell composition: creating a rich platform of composed services
  – Model-based Cell configuration and control
  – Management system automation across large-scale infrastructures

• Business impact: Cloud services
Service Lifecycle Management (SLiM)

- Configuring, deploying and managing cloud services
- Models and model transforms for the complete service lifecycle
  - From the ‘kit of possible parts’ through to the deployed service
- Multi-viewpoint, modular, parametric models
- Managing and predicting non-functional properties
- Lifecycle is round-trip, and includes service adaptation, service upgrade, etc.
- Delivered as a service (of course!)
HP/Intel/Yahoo! Global Cloud Computing Research Test-Bed*

- Codename *Cirrus*
- Sponsors:
  - HP Labs, Intel Research, Yahoo! Research
- Partners:
  - IDA Singapore
  - Karlsruhe Institute of Technology
  - University of Illinois at Urbana Champaign
- A context for collaborative research on management of Cloud Computing data-centres and applications
- A shared, large-scale, distributed test-bed
- Great coverage for the public announcement ... now getting started with the technical agenda

* Official Title
Thank you!
Why Cloud Computing?

- **Cost reduction**
  - Benefit from economies of scale
  - Predictability of spend
  - Avoids cost of over-provisioning
  - Reduction in up-front investment

- **Risk reduction**
  - Someone else worries about running the data-centre, protecting your data, and providing DR
  - Reduces risk of under-provisioning

- **Flexibility**
  - Add/remove use of services
  - Scale up and down as needed
  - Do it quickly

- **Service Evolution**
  - Services evolve and (hopefully!) improve behind the scenes, with no user-involvement required

- **Ubiquity**
  - access from any place, any device, any time
And ...

- At massive scale  
  - Millions of users
- With unprecedented flexibility  
  - Mash-ups, aggregation, enhancing services, flexing up and down, ...
- Offering evolving APIs to exploit and extend
- At breakthrough cost levels  
  - Economies of scale  
  - New revenue models  
  - Eliminating old sources of cost (SaaS vs. CD)
Approaches to reduce cost

I.T. “islands”
1:1 binding
Workloads: Machines

 Adaptive Infrastructure
“Pooled” resources/Shared infrastructure

Dedicated  Virtualized  Automated

Cost continuum

$X  $X/(2-3)  $X/10

Cloud
Narrowing the Chasm

All the workloads in the world

Cloud research test bed

Cloud Ready

Research Focus
Pain points for Users

- Experience spans multiple applications and multiple devices
  - Inconsistent behavior, inconsistent data, nothing where you need it
  - It’s easy to blame the devices
  - Even on the same device applications communicate and integrate poorly
  - Off-line changes, local copies: where’s the right data?
- User has to explain to each device and app what’s wanted
- User has to coddle and cajole the technology
Cloud Contribution to User Experience

- Many, many services collaborating
- Many services reasoning over data maintained by many other services

Reasoning services
- Coordinate your activities
- Compose other services to provide experiences
- Identify patterns, intentions
- Offer assistance, suggestion, opportunities
- Many more

Data maintaining services
- Know your location
- Know your immediate and long running activities
- Know your preferences
- Know your contacts
- Much more
Challenges for Developer

• Designing for
  – Performance at Scale
  – Redundancy
  – Multi-tenancy

• Development time & cost

• Accessing quality information from all devices & other services

• Knowing how customers use their services
Cloud Contribution to Developer Experiences

- Programming model to make it easy, fast, and cost effective to:
  - deliver services to the cloud, and,
  - consume services from the cloud
- Focus effort on unique value, delegate the rest
- Reduce testing and integration costs
- Access data from all sources
- Analytics to learn from customer usage
- Fast release cycles, easy to iterate, quick, agile
Mercado

- Enable “service marketplaces” where businesses can obtain services from a global ecosystem of service providers, rapidly, efficiently, and with high quality of service
- Research contributions
  - Business process specification, service selection, instantiation, choreography and adaptation over independently created service components
  - Service-oriented and model-based architectures, combined with Web 2.0, social networking, and semantic web mechanisms
Barriers to Enterprise Adoption
(aka Opportunities)

• Trust in the service vendor
  − Service levels
  − Stability
  − Geographic presence
• ISV support not widespread
• Few have taken the plunge in a big way
• Customizability of service offerings for specific needs of each enterprise
• Concerns about lock-in, lack of multi-vendor options
• Data: privacy, locality
• Regulatory concerns
• Challenge of migrating from in-house (or outsourced) apps
• Vested interests!
Inside the Cloud

Google in 2007
- 36 data centers containing > 800K servers
- 40 servers/rack
- Custom hardware, customized Linux base OS, lots of homebrew s/w
- Gobbling up dark fibre

- Many unreliable servers to fewer high cost servers
- Single search query touches 700 to up to 1k machines in < 0.25sec
- Typical H/W failures: Install 1000 machines and in 1 year you’ll see: 1000+ HD failures, 20 mini switch failures, 5 full switch failures, 1 PDU failure
- There are more than 200 Google File System clusters
- The largest BigTable instance manages about 6 petabytes of data spread across thousands of machines