云计算技术及应用

华清远见：郑老师

zm@farsight.com.cn

Gartner Report

- Top 10 Strategic Technology Areas for 2009
  - Virtualization
  - Cloud Computing
  - Servers: Beyond Blades
  - Web-Oriented Architectures
  - Enterprise Mashups
  - Specialized Systems
  - Social Software and Social Networking
  - Unified Communications
  - Business Intelligence
  - Green Information Technology
- Top 10 Strategic Technology Areas for 2010
  - Cloud Computing
  - Advanced Analytics
  - Client Computing
  - IT for Green
  - Reshaping the Data Center
  - Social Computing
  - Security – Activity Monitoring
  - Flash Memory
  - Virtualization for Availability
  - Mobile Applications

From Desktop/HPC/Grids to Internet Clouds in 30 Years

- HPC moving from centralized supercomputers to geographically distributed desktops, clusters, and grids to clouds over last 30 years
- R/D efforts on HPC, clusters, Grids, P2P, and virtual machines has laid the foundation of cloud computing that has been greatly advocated since 2007
- Location of computing infrastructure in areas with lower costs in hardware, software, datasets, space, and power requirements – moving from desktop computing to datacenter-based clouds

What is Cloud Computing?

- 1. Web-scale problems
- 2. Large data centers
- 3. Different models of computing
- 4. Highly-interactive Web applications

1. “Web-Scale” Problems

- Characteristics:
  - Definitely data-intensive
  - May also be processing intensive
- Examples:
  - Crawling, indexing, searching, mining the Web
  - Data warehouses
  - Sensor networks
  - “Post-genomics” life sciences research
  - Other scientific data (physics, astronomy, etc.)
  - Web 2.0 applications
  - …

How much data?

- Google processes 20 PB a day (2008)
- "all words ever spoken by human beings" ~ 5 EB
- CERN's LHC will generate 10-15 PB a year

640K ought to be enough for anybody.

What to do with more data?

- Answering factoid questions
  - Pattern matching on the Web
  - Works amazingly well
- Learning relations
  - Start with seed instances
  - Search for patterns on the Web
  - Using patterns to find more instances

How do I make money?

- Petabytes of valuable customer data…
  - Sitting idle in existing data warehouses
  - Overflowing out of existing data warehouses
  - Simply being thrown away
- Source of data:
  - OLTP
  - User behavior logs
  - Call-center logs
  - Web crawls, public datasets
  - …
- Structured data (today) vs. unstructured data (tomorrow)
- How can an organization derive value from all this data?

2. Large Data Centers

- Web-scale problems? Throw more machines at it!
- Centralization of resources in large data centers
  - Necessary ingredients: fiber, juice, and land
  - What do Oregon, Iceland, and abandoned mines have in common?
- Important Issues:
  - Efficiency
  - Redundancy
  - Utilization
  - Security
  - Management overhead

3. Different Computing Models

- Utility computing
  - Why buy machines when you can rent cycles?
  - Examples: Amazon’s EC2
- Platform as a Service (PaaS)
  - Give me nice API and take care of the implementation
  - Example: Google App Engine
- Software as a Service (SaaS)
  - Just run it for me!
  - Example: Gmail

“Why do it yourself if you can pay someone to do it for you?”

4. Web Applications

- What is the nature of future software applications?
  - From the desktop to the browser
  - SaaS == Web-based applications
  - Examples: Google Maps, Facebook
- How do we deliver highly-interactive Web-based applications?
  - AJAX (asynchronous JavaScript and XML)
  - A hack on top of a mistake built on sand, all held together by duct tape and chewing gum?

Some Cloud Definitions

- Ian Foster et al defined cloud computing as a large-scale distributed computing paradigm, that is driven by economics of scale, in which a pool of abstracted virtualized, dynamically-scalable, managed computing power, storage, platforms, and services are delivered on demand to external customers over the internet(云计算是一种商业计算模型。它将计算任务分布在大量计算机构成的资源池上，使各种应用系统能够根据需要获取计算力、存储空间和各种软件服务。)
- IBM experts consider clouds that can:
  - Host a variety of different workloads, including batch-style backend interactive, user-facing applications
  - Allow workloads to be deployed and scaled-out quickly through the rapid provisioning of virtual machines or physical machines
  - Support redundant, self-recovering, highly scalable programming models that allow workloads to recover from HW/SW failures
  - Monitor resource use in real time to rebalance allocations on demand

Internet Cloud Goals

- Sharing of peak-load capacity among a large pool of users, improving overall resource utilization
- Separation of infrastructure maintenance duties from domain-specific application development
- Major cloud applications include upgraded web services, distributed data storage, raw supercomputing, and access to specialized Grid, P2P, data-mining, and content networking services

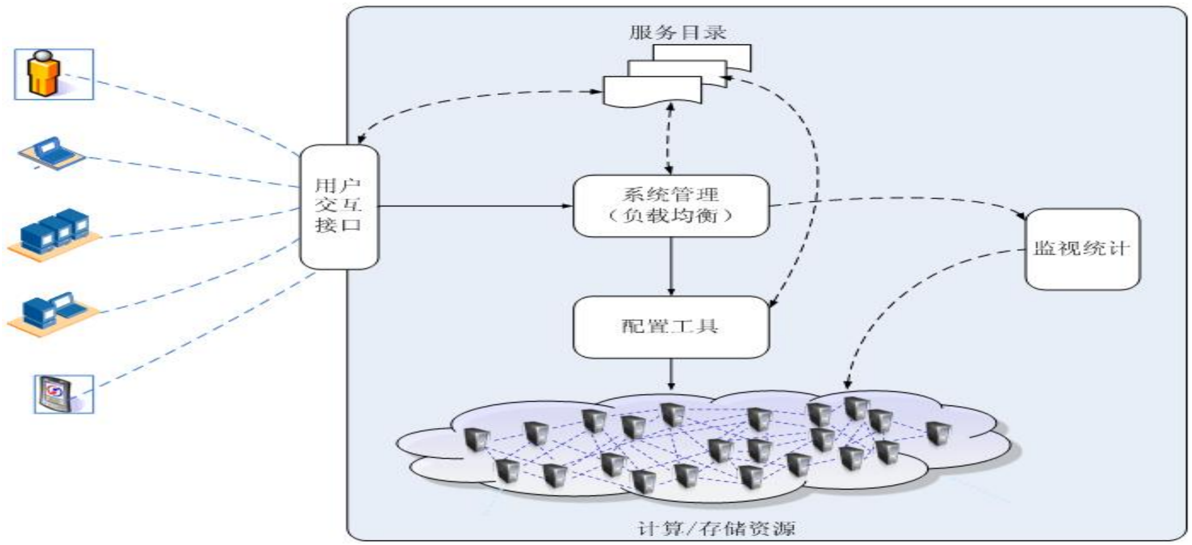
Three Aspects in Hardware that are New in Cloud Computing

- The illusion of infinite computing resources available on demand, thereby eliminating the need for cloud users to plan far ahead for provisioning
- The elimination of an up-front commitment by cloud users, thereby allowing companies to start small and increase hardware resources when needed
- The ability to pay computing resources on a short-term basis as needed (e.g., processors by the hour and storage by the day) and release them after done and thereby rewarding resource conservation

Some Innovative Cloud Services and Application Opportunities

- Smart and pervasive cloud applications for individuals, homes, communities, companies, and governments, etc.
- Coordinated Calendar, Itinerary, job management, events, and consumer record management (CRM) services
- Coordinated word processing, on-line presentations, web-based desktops, sharing on-line documents, datasets, photos, video, and databases, etc
- Deploy conventional cluster, grid, P2P, social networking applications in cloud environments, more cost-effectively
- Earthbound Applications that Demand Elasticity and Parallelism rather data movement Costs

Operations in Cloud Computing

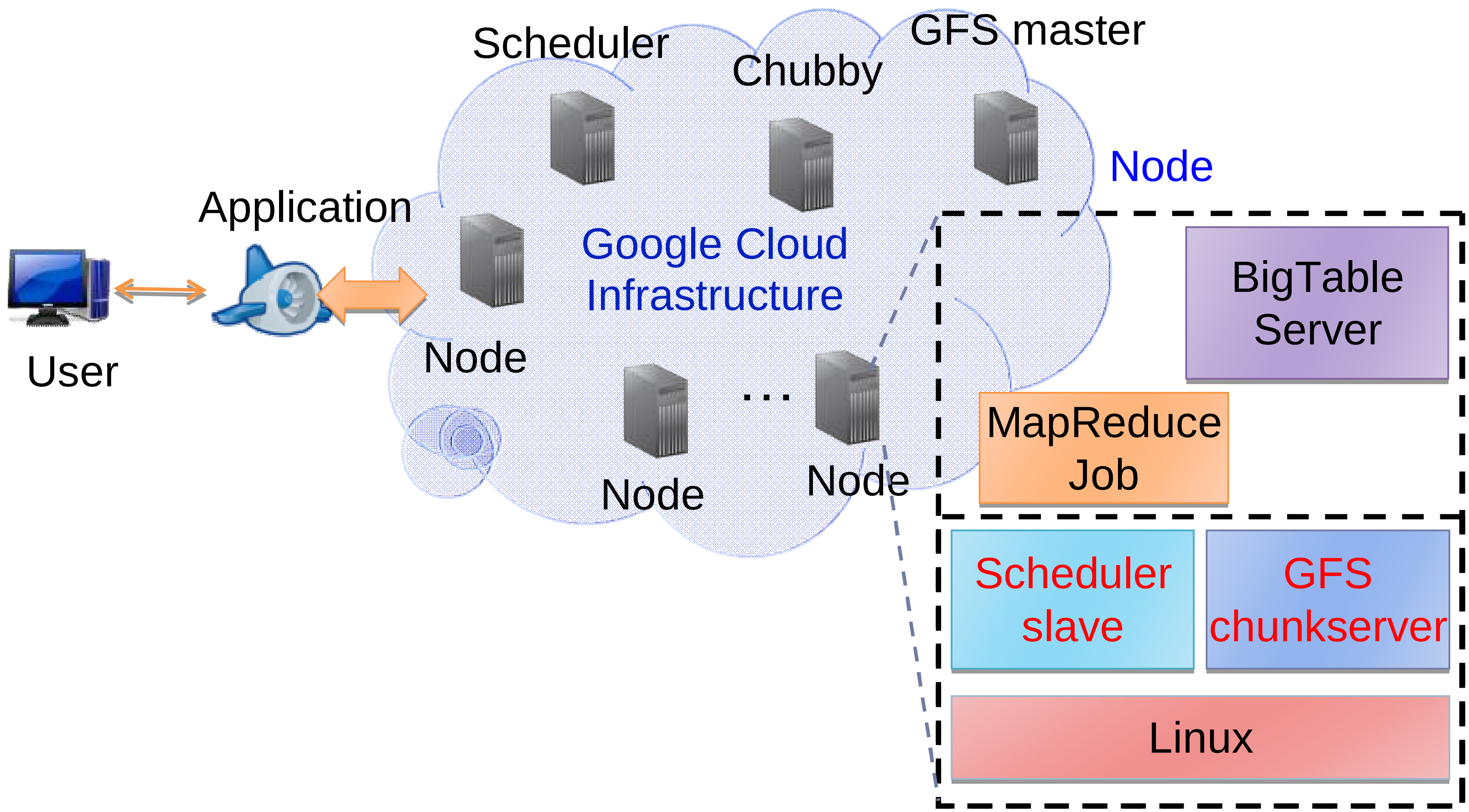


- Users interact with the cloud to request service
- Provisioning tool carves out the systems from the cloud – configuration or reconfiguration, or deprovision
- The servers can be either real or virtual machines
- Supporting resources include distributed storage system, datacenters, security devices, etc.

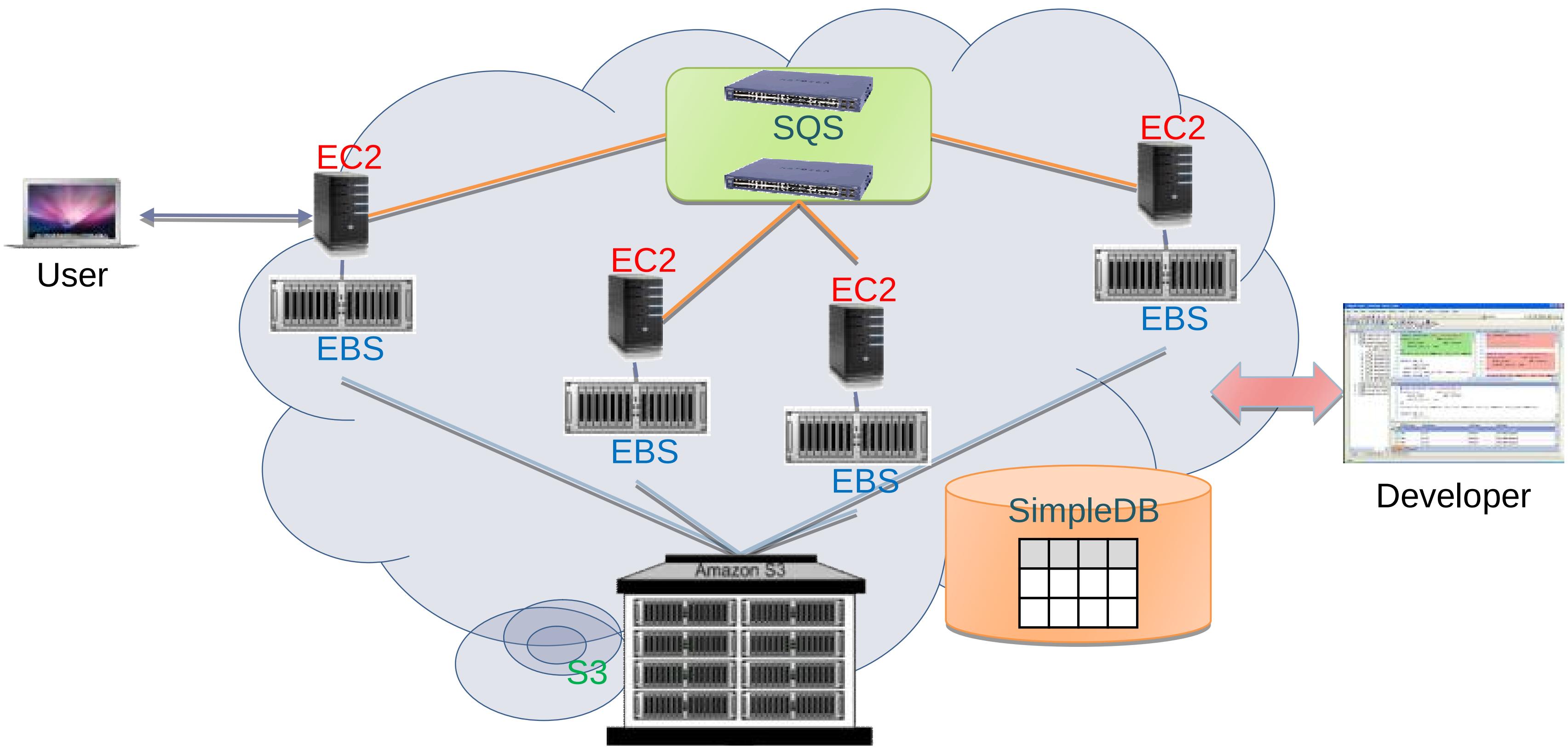
Cloud Computing Instances

- Google
- Amazon
- Microsoft Azure
- IBM Blue Cloud

Google Cloud Infrastructure

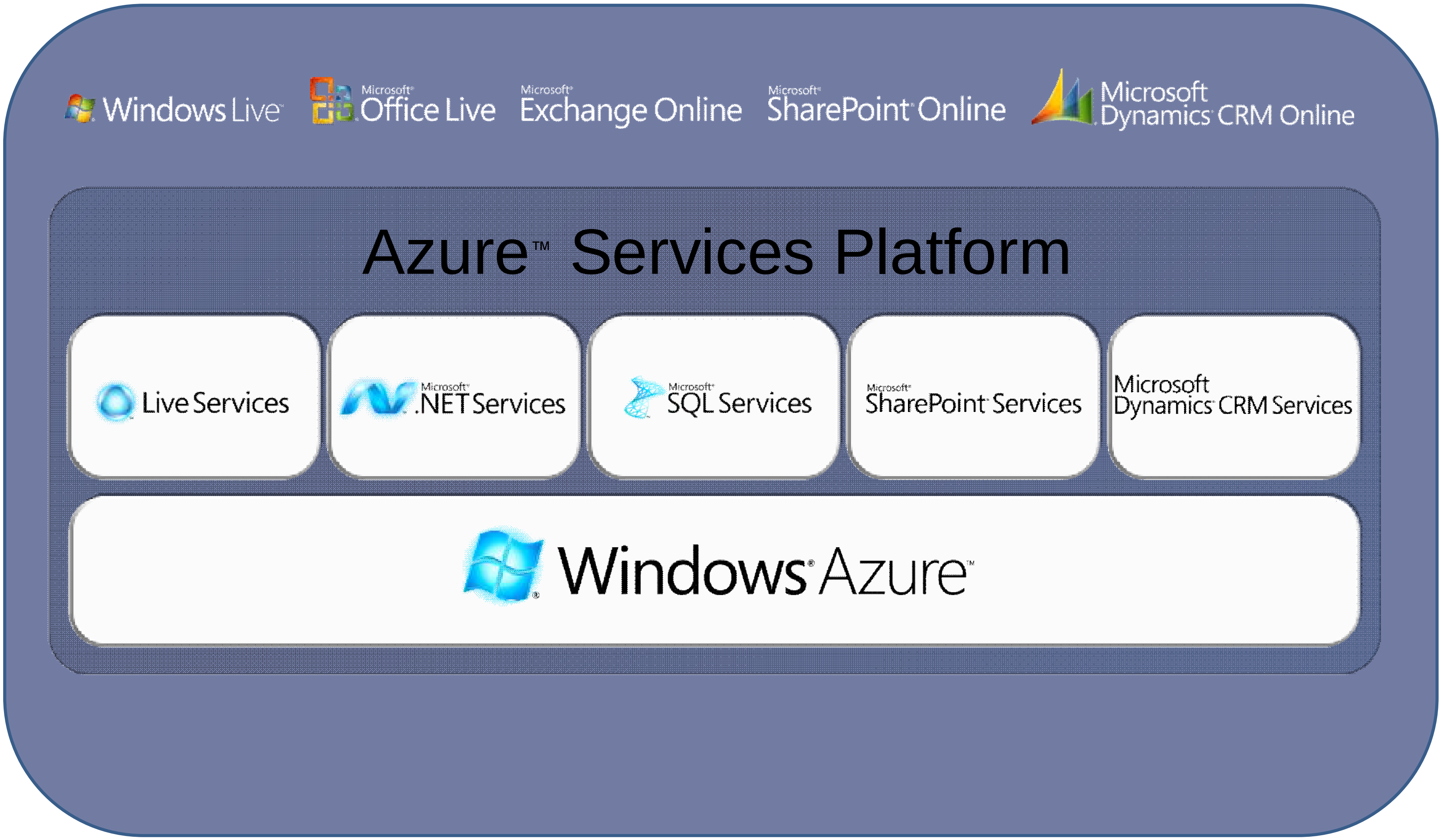


Amazon Elastic Computing Cloud

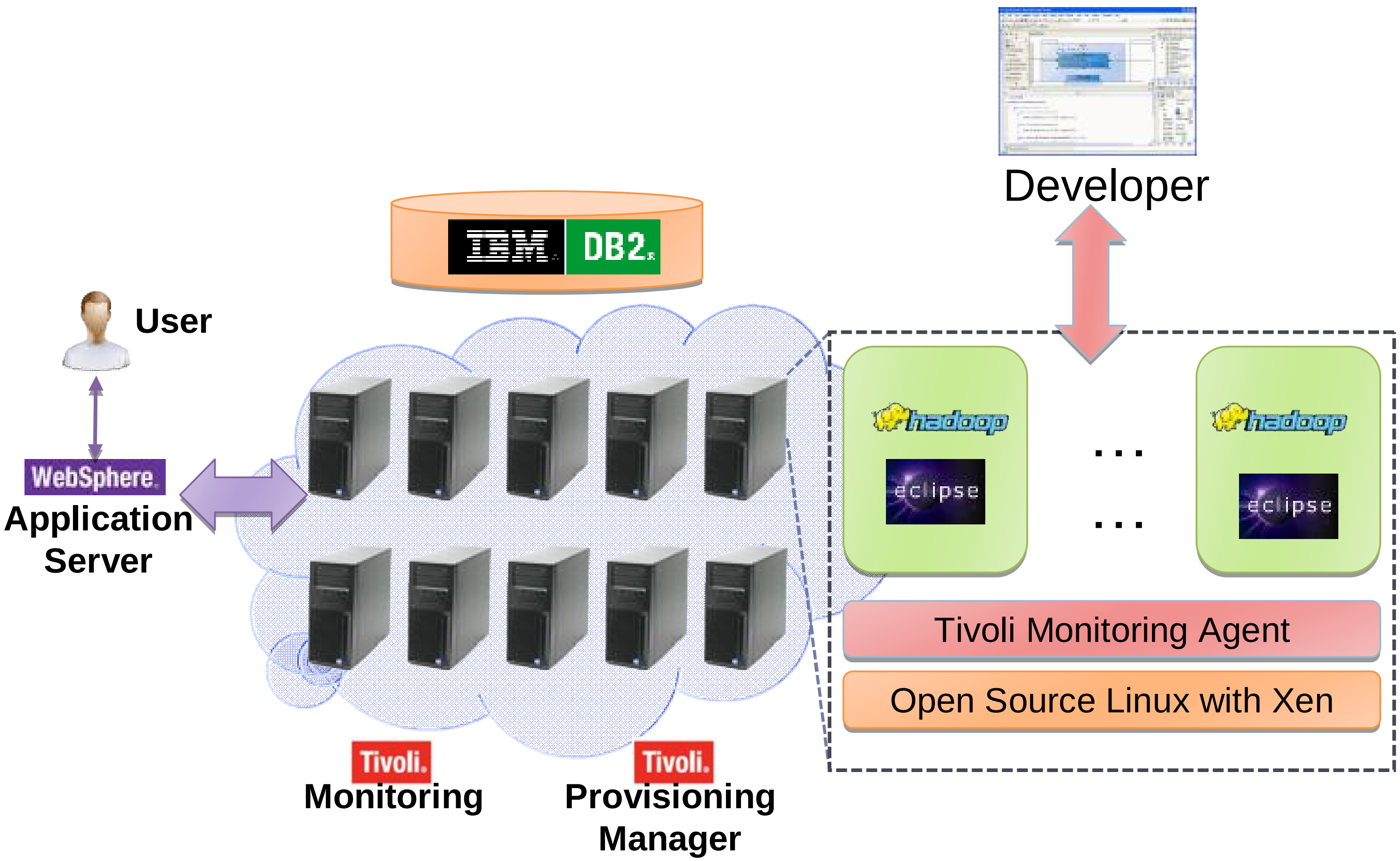


SQS: Simple Queue Service
EC2: Running Instance of Virtual Machines
EBS: Elastic Block Service, Providing the Block Interface, Storing Virtual Machine Images
S3: Simple Storage Service, SOAP, Object Interface
SimpleDB: Simplified Database

Microsoft Azure Platform



IBM Blue Cloud



Cost Considerations : Power, Cooling, Physical Plant, and Operational Costs

- Cost
  - technology costs
  - cost of security
  - etc.
- Benefits
  - availability
  - opportunity
  - consolidation
  - etc.

Cost Breakdown

- + Storage ($/MByte/year)
- + Computing ($/CPU Cycles)
- + Networking ($/bit)

Research Challenges

- Service availability
  - S3 outage: authentication service overload leading to unavailability
  - AppEngine partial outage
  - programming error
  - Gmail: site unavailable
  - Solutions:
    - The management of a Cloud Computing service by a single company results in a single point of failure (SPF).
    - In the Internet, a large ISP uses multiple network providers so that failure by a single company will not take them off the air. Similarly, we need multiple Cloud Computing providers to support each other to eliminate SPF.

Research Challenges

- Data Security
  - Current cloud offerings are essentially public rather than private networks, exposing the system to more attacks such as DDoS attacks.
  - Solutions:
    - There are many well understood technologies such as encrypted storage, virtual local area networks, and network middle boxes.

Research Challenges

- Data Transfer Bottlenecks
  - **Applications continue to become more data-intensive. If we assume applications may be "pulled apart" across the boundaries of clouds, this may complicate data placement and transport.**
  - **Both WAN bandwidth and intra-cloud networking technology are performance bottleneck.**
- **Industrial solutions:**
  - **It is estimated that 2/3 of the cost of WAN bandwidth is consumed by high-end routers, whereas only 1/3 charged by fiber industry.**
  - **We can lower the cost by using simpler routers built from commodity components with centralized control, but research is heading towards using high-end distributed routers .**

Research Challenges

- Software Licensing
  - Current software licenses commonly restrict the computers on which the software can run. Users pay for the software and then pay an annual maintenance fee.
  - Many cloud computing providers originally relied on open source software in part because the licensing model for commercial software is not a good match to Utility Computing.
- Some ideas:
  - We can encourage sales forces of software companies to sell products into Cloud Computing.
  - Or they can implement pay-per-use model to the software to adapt to a cloud environment.

Research Challenges

- Scalable storage
  - Differences between common storage and cloud storage
    - The system is built from many inexpensive commodity components that often fail
    - The system stores a modest number of large files
    - The workloads primarily consist both large streaming reads and small random reads. The workloads many large, sequential writes that append data to files and once written, files are seldom modified again.
  - The cloud storage (file) system needs to share many of the same goals as previous distributed file systems such as performance, scalability, reliability, and availability.
  - In addition, its design needs to be driven by key observations of the specific workloads and technological environment, both current and anticipated, that reflect a marked departure from some earlier file system design assumptions.
  - GFS
    - Files are divided into fixed-size chunks, Chunk size is one of the key design parameters. GFS chooses 64 MB, which is much larger than typical file system block sizes.
    - The master stores three major types of metadata: the file and chunk namespaces, the mapping from files to chunks, and the locations of each chunk's replicas.
    - GFS supports the usual operations to create, delete, open, close, read, and write files.

Research Challenges

- Transparent Programming Model
  - Programs written for cloud implementation need to be automatically parallelized and executed on a large cluster of commodity machines.
  - The run-time system should take care of the details of partitioning the input data, scheduling the program's execution across a set of machines, handling machine failures, and managing the required inter-machine communication.
  - The programming model should allow programmers without many experiences with parallel and distributed systems to easily utilize the resources of a large distributed system.
  - MapReduce
    - Scalable Data Processing on Large Clusters
    - A web programming model implemented for fast processing and generating large datasets
    - Applied mainly in web-scale search and cloud computing applications
    - Users specify a map function to generate a set of intermediate key/value pairs
    - Users use a reduce function to merge all intermediate values with the same intermediate key.

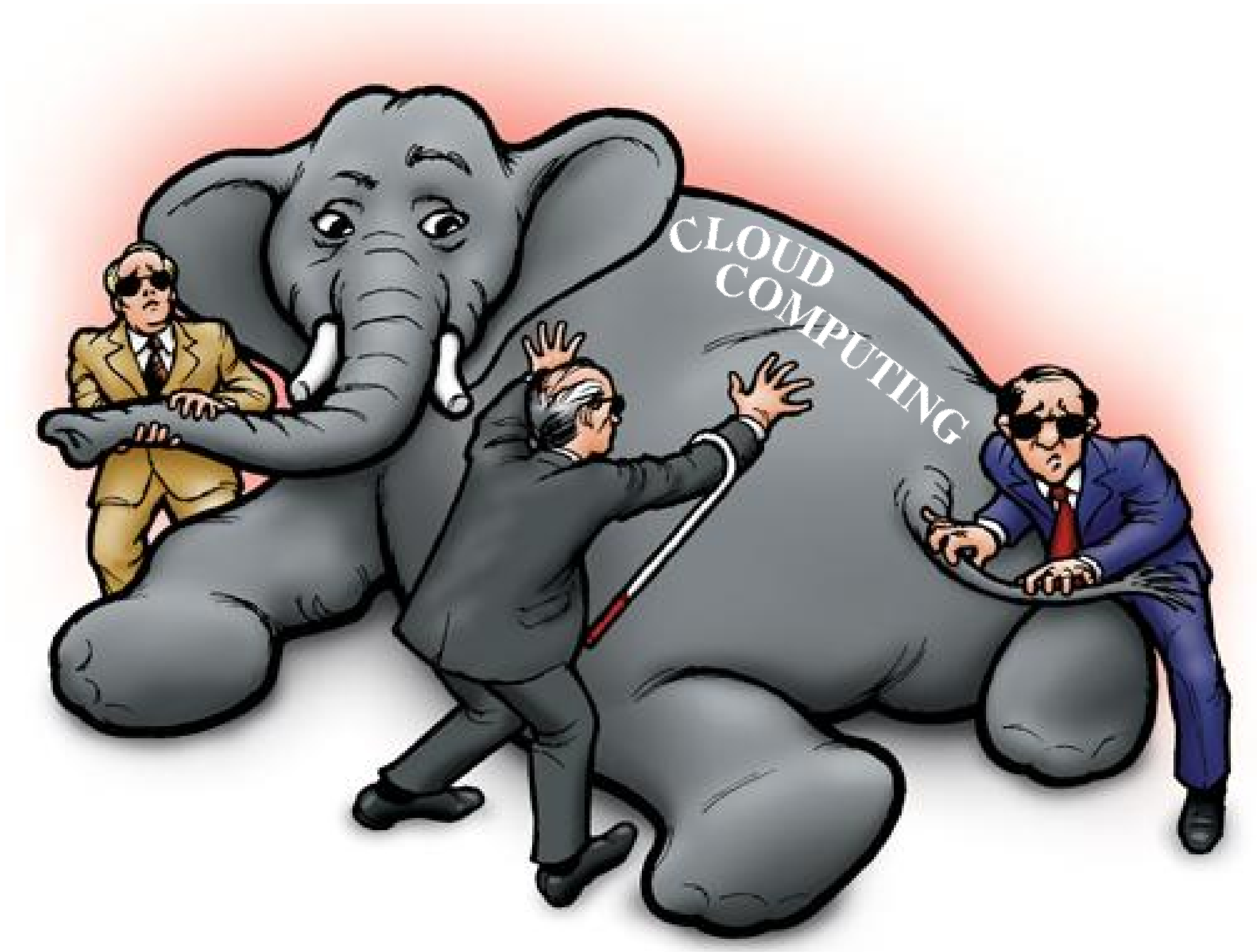
Research Challenges

- …

Steve Ballmer's View on the Future of Cloud

- **Cloud creates opportunities and responsibilities**
- **Cloud learns and helps you learn, decide and take action**
- **Cloud enhances social and professional interactions**
- **The cloud wants smarter devices**
- **Cloud drives server advances that, in turn, drive the cloud**

Cloud Computing Skepticism



“Cloud computing is simply a buzzword used to repackage grid computing and utility computing, both of which have existed for decades.”

whatis.com

Definition of Cloud Computing

"The interesting thing about cloud computing is that we've redefined cloud computing to include everything that we already do. […]
The computer industry is the only industry that is more fashion-driven than women's fashion.
Maybe I'm an idiot, but I have no idea what anyone is talking about. What is it? It's complete gibberish. It's insane. When is this idiocy going to stop?"

Larry Ellison

Q&A