Cloud Computing

Francis C.M. Lau, HKU

"The sun always shines above the clouds."

- Paul F. Davis

BIG DATA **and Cloud**

- We embrace cloud not just because we need to process data
- Also because we need a platform (PaaS), certain software (SaaS), or hardware resources (IaaS)
- But true, BIG DATA made cloud happen a lot more quickly
  - You don't want to operate a power plant at home just to control a power-thirsty appliance

Cloud as Utility

"The long dreamed vision of computing as a utility is finally emerging." [Armbrust et al.]

- You plug in (the outlet) and play [but sometimes it won't]
- You thought it is an infinite power source [but sometimes it'd run low, or even run out; and more often, it behaves unstably]
- You assume it is "elastic" – you use what you need exactly and pay for just that [but sometimes it won't stretch, sometimes it breaks, and you're charged unfairly]
- You thought everything is pretty safe [but didn't realize it could be a black hole]

Subtopic: Service Availability

- Dropbox "dropped out" on Jan. 10, 2014 for 2 days
- Clouds are a huge assemblage of components, and software has bugs!
- If your server at home hangs, you reboot, but you can't when a cloud hangs
- Distributed Denial of Service (DDoS) attacks are real
- **RQ: How to design a cloud service that is highly available?**
- **RQ: How to counter the "attacks"?**
- **RQ: How to tolerate faults or failures of components?**

Subtopic: Performance Predictability

- Fact: most virtualized environments have highly variable performance
- Variance also due to multi-tenancy, movements of large amounts of data, and the system itself (e.g., HDFS randomly distributes data blocks across a cluster)
- Even if CPU and memory sharing is not a problem, I/O sharing could easily kill performance
- Many HPC applications need to ensure that all the threads of a program are running simultaneously
- **RQ: How to make performance more predictable?**
- **RQ: How to guarantee performance/QoS?**

Subtopic: Providing Elasticity

- Scalability is key: quick, automatic scale up or down according to user's changing needs
- Application's scalability is another issue
  - 1 machine x 100 hrs = 100 machines x 1 hr?
- Ideally, you pay as you go, and are charged by the cycles (compute), or the bytes (storage and communication)
- **RQ: How to predict and react to workload changes quickly and dynamically?**
- **RQ: How to reduce bottlenecks and provide for the best speedups?**
- **RQ: How to charge more accurately and fairly?**
- **RQ: How to scale data storages?**

Subtopic: Data Confidentiality

"The main issue is that expectations of trustworthiness may be unrealistic." [Neumann]

- Apparently there should be no "fundamental" obstacles to making a cloud-computing environment as secure as in-house IT environments
  - But clouds do have a lot more weak spots
- Gartner: 50% of enterprises will use hybrid cloud (which includes a private cloud) by 2017
  - Also for performance reasons: some data are "earthly"
- **RQ: How to make cloud sufficiently secure and trustworthy?**

Subtopic: Data Lock-In

- Although software stacks have improved interoperability among platforms, APIs for cloud applications are still predominantly proprietary
- Customers cannot easily extract their data and programs from one site to run on another
- It is really "vendor lock-in"
- **RQ: Standardization of APIs?**
- **RQ: How to design a heterogeneous cloud that would integrate parts from multiple vendors?**

Subtopic: Optimizing Data Placement and Transfer

- Big Data: applications easily get "pulled apart" across the boundaries of machines or even clouds
- Cost and performance depend a lot on data placement and transport
  - Jim Gray: The cheapest way to send a lot of data is to physically send disks or even whole computers via overnight delivery services
- **RQ: How to place and re-place data such that the best cost-performance can be achieved?**

THE CLOUD BEGINS WITH COAL

BIG DATA, BIG NETWORKS, BIG INFRASTRUCTURE, AND BIG POWER

AN OVERVIEW OF THE ELECTRICITY USED BY THE GLOBAL DIGITAL ECOSYSTEM

Mark P. Mills
CEO, Digital Power Group
www.tech-pundit.com

August 2013

"The ICT ecosystem (the Internet, Big Data, and the Cloud) now approaches 10% of world electricity generation"

- Amazon: energy-related costs: 42% of total (19% power; 23% cooling) [2009] (now much improved)
- Cloud computing (due to server consolidation) is considered green computing, but the computers they use may not be green

Subtopic: Green Cloud

- Existing solutions: Energy efficient hardware, processor-level energy-aware scheduling (e.g., DVS)
- Even when run at a low utilization, servers typically need up to 70% of their maximum power consumption
- Virtualization increases energy efficiency
- **RQ: How to perform energy-aware scheduling?**
- **RQ: How to achieve the best tradeoff in computation/communication/storage and energy/performance?**

Emerging Opportunities

- Thin interactive apps that are backed by the cloud, even when they are disconnected
  - Mobile cloud
  - Edge computing, fog computing
- Cloud and IoT
  - Most "things" are not computers
- Data intensive batch processing for business analytics
  - Less online transactions, more decision support
- Compute-intensive desktop apps
  - Symbolic math, 3D rendering, …

More RQs by Colleagues

- Cloud accesses are remote and have low performance. ***Caching*** improves performance but is subject to reliability challenges. How to design high-performance and high-persistent caching strategies?
- Integrating multiple clouds (***cloud-of-clouds***) can boost scalability, but how to address the heterogeneity of different clouds?
- How to design ***dynamic pricing mechanisms*** that are optimal?
- How to support ***online education*** and ***remote health*** through a cloud platform?
- How to jointly optimize network and data resources in order to achieve effective ***geo-diversity*** in datacenter design?

… Hong Kong

- Ideal location for datacenters, data hub
  - Cf. the "Enhancing Hong Kong's strategic position as a regional and international business center" theme
- Green cloud
  - Cf. the "Developing a sustainable environment" theme
- Mobile cloud
  - HK ranks #1 by connections/citizen (March 2015)
- Adoption by SMEs and startups
  - "It used to take years to grow a business to several million customers – now it can happen in months." [Armburst et al.]
- We're very strong in Data Engineering, Networking, Cloud, …

References

- Michael Armbrust et al., "A View of Cloud Computing", *CACM*, Volume 53 Issue 4, April 2010.
- Andreas Berl, Erol Gelenbe, Marco di Girolamo, Giovanni Giuliani, Hermann de Meer, Minh Quan Dang, and Kostas Pentikousis, "Energy-Efficient Cloud Computing", *The Computer Journal*, Vol. 53 No. 7, 2010.
- Ken Birman, Gregory Chockler, and Robbert van Renesse, "Toward a cloud computing research agenda", *ACM SIGACT News*, Volume 40 Issue 2, June 2009.
- Peter G. Neumann, "Inside Risks Risks and Myths of Cloud Computing and Cloud Storage", *CACM*, Volume 57 Issue 10, October 2014.
- Malte Schwarzkopf, Derek G. Murray, and Steven Hand, "The Seven Deadly Sins of Cloud Computing Research", *HotCloud 2012*.