Cloud Computing Research Challenges

Johan Tordsson
The overall research problem:

How much and what type of resources to allocate and when and where to deploy them?
Why is this important, i.e., who cares?
Users care (about slow or unreachable online services…)

What would you do...
• if it would take 5 seconds to view each hotel at an online booking site?
• if the web session crashes during payment transactions?

Perhaps as everyone else...
• 82% of customers give up on failed payment*
• 25% of users disappear if page load time > 4 s**
• 1% lost sales for each 100 ms delayed response time**
Google: 0.5 s slower page lost → 20% lost revenue ***

Service providers care about insufficient server capacity!

* JupiterResearch  ** Amazon  ***Google
Infrastructure providers care about data center efficiency

- **Power usage escalation**
  - Datacenters - more and larger every year, estimated to use 3% of global energy production
  - Facebook in Luleå will use power equivalent to heating for 40’000 homes
  - ...

- **Utilization is poor**
  - CPU and memory utilization barely half of available capacity
  - 84% of VMs reach max utilization less 20% of time
  - ...

UMU group research overview
Relation to other research fields

Collaborations

Related research communities
Challenge: industrial competition

• Clouds were invented by industry
  • A research field originating in how an online bookstore handled its servers...

• Industry (research) leader in cloud
  • Access to large-scale infrastructures
    • Academic “datacenters” are not

• Access to real workloads/datasets
  • Academia is stuck mostly with old and/or irrelevant traces (FIFA’98 etc.)

• Access to real applications
  • Small set of publicly available benchmarks

• Scale and budget
  • Google, Amazon, Microsoft, etc. can solve the same problem faster
Academic Opportunities

• Freedom & long-term perspective
  • With good funding, academia can focus 5+ years forward

• Industry has its hand full with:
  • Operation of their existing services and infrastructure “changing engines on a flying airplane…”, 5-10 years…
  • Quartile economics

• When industry is (too) interested in your research, you are doing the wrong thing
  • Avoid time-to-market races…

• Aim for the “that’s really cool, but I have no idea how to do that” problems

• Ambition and risk?
  • In academia, it is more ok to fail, but the occasional success will have large impact
  • Industry expects (reasonable) success
Academic Opportunities (cont.)

• (Safe) bets on future technologies/environments
  
  • No need to wait for technological development to happen
    • e.g., SSDs were anticipated long before first built
  
  • Which current limitations will (likely) go away?
    • Disk latency, network bandwidth, speed of light?
    • What new research challenges will arise then?
  
  • Current development (too late to bet?)
    • Storage will become irrelevant
    • RAIN – Redundant Array of Inexpensive Nodes,
      keeping stuff in memory in 3 datacenters is very redundant...

• Sample bets by the UMU group:
  • Convergence of cloud and telco infrastructures
  • Virtually large machines seemingly spanning multiple servers
Collaborations?

• Why collaborate with industry?
  • Formulate more interesting problems
  • Broader expertise to solve problems
  • Positioning/branding towards future students, current students, funders, collaboration partners
  • Better papers (at better conferences)

• (European) challenges
  • Most relevant companies are US-based
  • Collaborating mostly with top US universities
    • Sending PhD students as summer interns
    • To later co-authoring papers during semesters
Conclusions

• Cloud computing is a high-impact, high-interest (research) area

• Industry research leader

• Academic research needs careful positioning

• Collaborations (for the right reasons) a way forward