1.264 Lecture 31

System architecture
Cloud computing

Next class: Green chapter 1-3. Exercise due before class
System architecture and configuration

• Organization of all system components (hardware, software, network) is the **system architecture**
• Done early in system development or configuration project
  – Assess users, applications, system software, networks, hardware
  – Configure Web/app/db servers, networks, backup, etc.
  – Configurations are complex and changing
  – Information is usually wrong on which estimates are based
  – Step is needed because you need a **budget** for a project
• Remember the estimate convergence curve
  – There is much uncertainty in the final system hardware and software. We are only in spiral 1 still.
  – If you make a baseline (point) estimate, you have a lot of error and about a 50% chance of being too low
  – If you’ve done everything else right but your system is slow because of inadequate hardware or telecom, it will be very frustrating
  – Your customer can’t tell why your system is slow. It may be bad database, bad/inefficient software implementation, etc.
  – This is one of the few problems that you can throw money at to solve!
• **Successful configurations are usually overbuilt**
  – Most components will be oversized, but you have a better chance of having enough capacity at the (unanticipated) bottleneck
  – Spend your entire budget, always, or use cloud computing

• **Each server (often virtualized) does just one thing:**
  – HTTP server (Web server)
  – Business logic (app server)
  – SQL (database server)
  – We can understand and characterize its task this way. If a box handles many functions, sizing and managing it is hard
Performance metrics

• **Metrics (database server as an example)**
  – **Throughput**: I/O operations/second, data transfer rate/second
    • This is **system** view
    • If throughput is low, more servers used than really needed
  – **Latency**: seek time (seconds), response time (seconds)
    • This is **user** view
    • If latency is high, users have slow response
  – **Utilization**: percentage of data transfer rate, disk capacity used
    • This is **future** view
    • If avg utilization > 60%, your systems will crawl on busy days
    • Average server utilization ~15-20% in traditional data centers

• **Units**
  – Ethernet data transfer: 100 Mbit/sec = 12.50 MB/sec
  – Disk data transfer: 10.00 MB/sec
  – Modem data transfer: 56,000 bits/sec = 0.007 MB/sec

• **Use comparable units when assessing systems**
  – 8 bits (b)= 1 byte (B)
  – MB may be 1,000,000 B or 1,024,000 B. GB, TB discrepancies higher.
Exercise 1: Database server cost, performance

• Log on to www.tpc.org
  – Results -> TPC-C -> Top 10 Price/Performance
  – Click on “system” for top performing choice and open executive summary as pdf. Pick the 2nd one: IBM
  – Look at the configuration: (Note the 50% discount)
    • What does the server hardware cost? How many users?
    • What do the disks cost? How many are there? Total storage?
    • What does the software cost? (This one is unusual.)
    • How many CPUs? How much memory?
    • Ignore the client hardware, which is just for the test
    • Look at the range of performance (response times)
  – Results -> TPC-H -> Top 10 Price/Performance
    • Look at 100GB results, same items as TPC-C.
    • The units are queries per hour.
    • Compare with TPC-C performance, which is transactions per minute. Why the difference?
Solution (November 2013)

- **TPC-C: IBM x3650**
  - What does the server cost?  ~$30,000. Users: ~1,000,000
  - What do the disks cost?  ~$260,000
  - Software:  ~$250,000
  - How many disks are there?  115, holding 40,000GB
  - How many CPUs?  2 (8 cores)
  - How much memory?  768GB
  - Look at the range of performance (response times) ~0.1 seconds

- **TPC-H: Lenovo ThinkServer RD630 (100 GB database)**
  - What does the server cost?  ~$13,500. Users: a few (5-10)
  - What do the disks cost?  (not specified, but most of it)
  - How many disks are there?  8, holding 2,400GB
  - How many CPUs?  2 (16 cores)
  - How much memory?  64GB
  - Look at the range of response times: 0.3-12.7 seconds
  - Compare: 420,000 QpH= 7,000 Q/min(H) vs 1,300,000 q/min(C)
    - H transaction is about 1,000,000 times the cost of a C transaction
Servers are faster/cheaper now

• Best 2000 TPC-C (price/performance):
  – 8,000 tpmC, $58/tpmC, 6,000 users, 5 second response

• Best 2013 TPC-C (price/performance):
  – 1,600,000 tpmC, $0.47/tpmC, 1,000,000 users, 0.1 second response
  – ~3,200,000 total transactions/minute in benchmark
  – ~800,000 total transactions/minute actual max
    • 60% utilization max for good performance
    • Your system is less highly tuned than benchmark app
    • Peaking
    • Allow for future growth
  – Mix of transactions changing: more status/inquiry with Web services

• Highest performance 2013:
  – 30,000,000 tpmC, $1/tpmC, 24,000,000 users
Cloud computing vs traditional options

• Amazon, Microsoft Azure, Google, others
  – No need for Web/application/database servers
  – No need for Web/database software
  – Write your application and place it in the cloud
    • Private or public
  – Pay by the hour

• Options for applications
  – Cloud computing
  – TPC/C-E-H systems that you purchase and host in-house or in a hosting center
  – Cheap servers from Dell, etc. that you configure for latency, throughput, utilization, etc. using rules
Cloud computing

• Three types of cloud:
  – **Infrastructure** as a service: outsource hardware
    • Provide own operating system, database, app, etc.
    • Most flexible, most expensive total IT cost
    • Examples: Rackspace
  – **Platform** as a service: outsource operating environment
    • Cloud provides OS, database, etc.
    • Example: Amazon, Microsoft, Google
  – **Software** as a service: no app of your own
    • Configure the third party app
    • Examples: salesforce.com
    • Least flexible, least expensive total IT cost
    • Transportation examples:
      – Active On-Demand APT (package carrier TMS)
      – One Network (logistics)
Exercise 2: Cloud apps

• Which cloud type, if any, would serve these apps?
  – Test and preproduction systems
  – Application, DB development tools (UML, Visual Studio, ...)
  – Batch processes needing limited security (e.g., raw material management)
  – Off-the-shelf software (e.g., HR, CRM, collaboration)
  – Multiple co-dependent apps
  – Apps with strict licensing and accountability requirements
  – Apps that require a lot of customization and flexibility

• Use the following abbreviations
  – IaaS: infrastructure as a service
  – PaaS: platform as a service
  – SaaS: software as a service
Solution: Cloud apps

- Which cloud type, if any, would serve these apps?
  - Test and preproduction systems **IaaS, PaaS**
  - Application, database development tools (UML…) software **SaaS**
  - Batch processes needing limited security (e.g., raw material management) **IaaS, PaaS**
  - Off-the-shelf software (e.g., HR, CRM, collaboration) **SaaS**
  - Multiple co-dependent apps **Not in cloud**
  - Apps with strict licensing and accountability requirements **Not in cloud**
  - Apps that require a lot of customization and flexibility **Not in cloud, or perhaps IaaS**
Exercise 3: cloud computing

• Visit these two Web sites to get prices: (What kind of cloud?)
  – aws.amazon.com/ec2/ Estimate on-demand and reserved:
    • Go to “EC2 Pricing”, “AWS Simple Monthly Calculator”, US East
    • Windows/Web SQL Svr, 50% utilization, m3.xlarge (15 GB memory, 13 CPU units) On demand and then reserved
    • Storage: 1000GB standard, 100 iops, 100GB-mo snapshot
    • Data: 1,000GB/month inter-region in and out
    • Reserved: Medium util, 1 yr term
    • 1 site, 150GB database (less than Amazon), 1000GB data transfer

• Glance at developers.google.com/appengine/
  – Go to “Quotas and pricing”->”Pricing”
  – Google costs require more work to estimate…

• Think about comparing to running your own data center:
  – Hardware, software, space, security, staff to manage, ….
Solution (November 2013)

- **Amazon cost:**
  - On demand: $488/month (was $831/month last year, for less)
  - Reserved: $1100 + $286/month (was $1837 + $590/month)
- **Microsoft cost:**
  - Shared: $355 PAYG, $275 12 mos (was $943/month last year)
- **Google:**
  - About the same…
System architecture summary

• Cloud computing growing exponentially
  – Cost, flexibility, higher quality software

• Major vendors offer cloud storage, computing, office applications, enterprise applications (CRM, SCM, ERP, etc.)
  – Software as a service (SaaS) is cloud computing
  – So are infrastructure and platform as a service (IaaS, PaaS)
  – Cloud computing is heavily based on Service Oriented Architectures (SOA)
    – Which are based on Web services
    – Which are based on HTTP and XML
    – Which exchange data between databases using SQL

• Private and public clouds exist

• Enterprise servers and applications will continue

• System architecture provides a cost estimate for the software and hardware
  – Typically required early in a project. Use first spiral to develop.
  – Estimate staff and software costs from lecture 3,4 approach.
System architecture examples

• Public transportation
  – Transit trip planner
  – Transit real time information
  – Transit fare payment
  – Automated vehicle location
  – Paratransit dispatch and control

• Supply chain
  – Warehouse management
  – Transportation management
  – Production management

• In all examples, architecture generates budget:
  – Staff: from function points, person months
  – Hardware: from TPC and cloud computing estimates
  – Telecom: from bandwidth and technology type
MIT OpenCourseWare
http://ocw.mit.edu

1.264J / ESD.264J Database, Internet, and Systems Integration Technologies
Fall 2013

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.