IS ARCHITECTURE OVERVIEW

2110684 Information System Architecture
Natawut Nupairoj, Ph.D.

Course Plan

- Overview and Background Concepts
- Networking and Web-Based Technology
- System Management / Security / Tuning

Size of Information

- How much computer storage is needed?
  - Bits, bytes, kilobytes, megabyte, terabyte, petabytes
- Examples:
  - Library of Congress: 3 PB
    - 20 TB for Books (20M books)
    - 13 TB for Photographs (13M photographs)
    - 200 TB for Maps (4M maps)
    - 500 TB for Movies (500K Movies)
    - 2,000 TB for CDs (3.5M sound recordings)
  - NASA Satellite Image DL: collection of 2 TB/day

Most common information today?

- Data records
  - Structured data
- Text, web pages, documents
  - Unstructured, or semi-structured data
- Images, video, music, voice,….
  - Multimedia data, multimedia documents
- Spatial/geographic data
  - Maps, spatial analysis data, census data, etc.

Tier/Logic Definition

- Presentation Logic: User Interfaces.
- Business Logic: Application Programs and Services.
- Data Logic: Databases.

Example: Registration System

- Presentation Logic: web pages
  - Login / Logout.
  - Register classes.
  - Add / Drop classes.
  - Change section.
  - Personal information.

Example: Registration System

- Business Logic: Modules and Services
  - Authenticate.
  - Add a class.
  - Drop a class.
  - Change section.
  - Show classes, grades, etc.
  - **Validate class selection (called by add/drop class).**

Example: Registration System

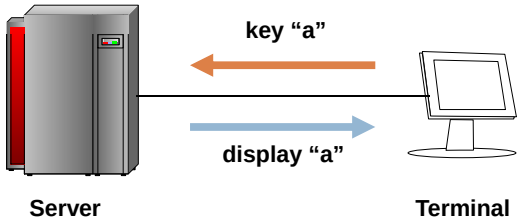
- Data Logic
  - Class database
    - Course number, title, credit, pre-requisite, etc.
  - Class availability database
    - Year, semester, course number.
  - Student database
    - Student-id, password, name, year.
  - Student class database
    - Year, semester, student-id, course number, grade.

Architectural Pattern: Transact

- Applications with business data is stored for a long time e.g. customer orders, student registration, etc.
- Types of architectures
  - 1-Tier.
  - 2-Tier (commonly known as client/server).
  - 3/N-Tier (commonly known as web-based).

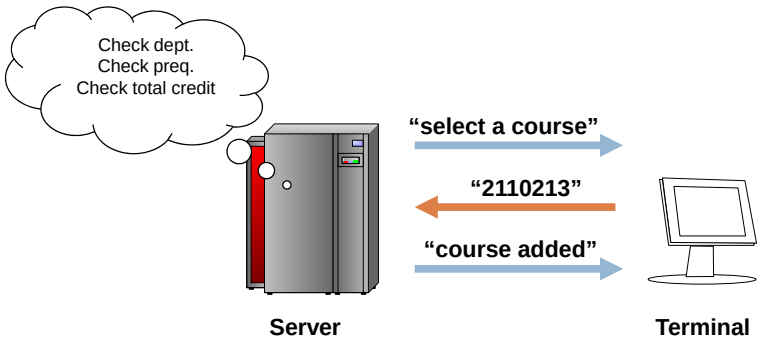
1-Tier Transact Pattern

- Server with "dumb" terminals.

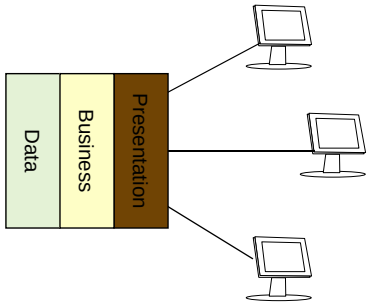


1-Tier Registration System

- Add a course:



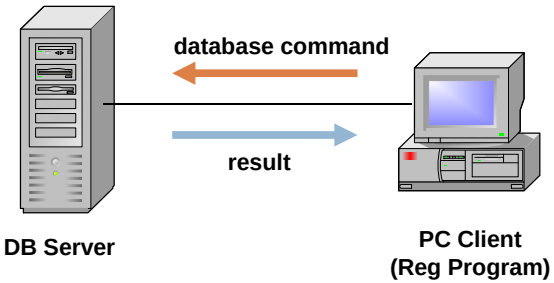
1-Tier Transact Pattern



- Pros:
  - Single place, easy to control.
- Cons:
  - Expensive, big spaghetti bowl, inflexible, lack of applications, take long time to develop.

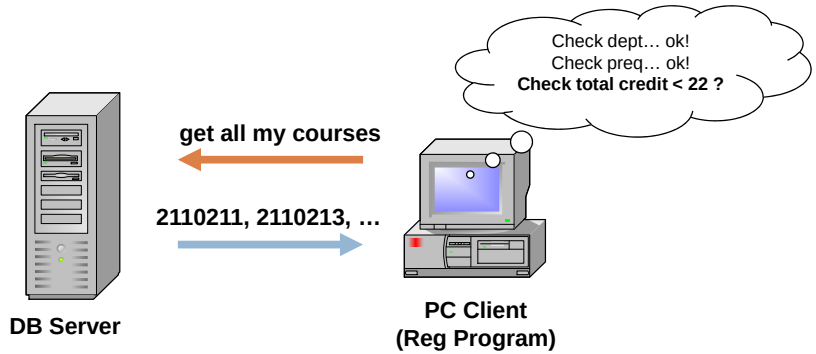
2-Tier Transact Pattern

- Database server with fat clients
  - VB / Developer 2000 client program + SQL database.

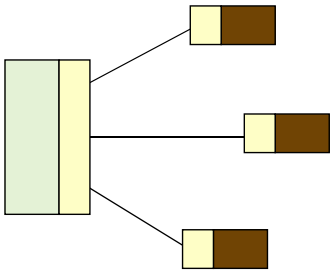


2-Tier Registration System

- Add a course:



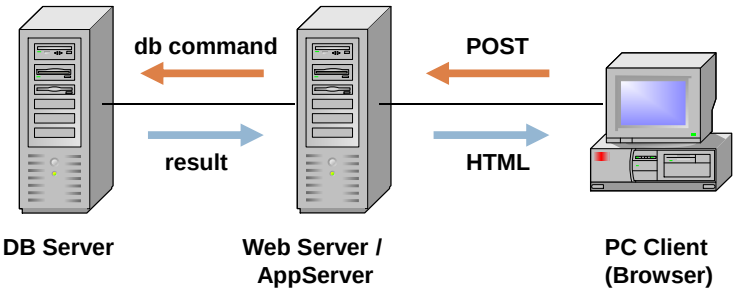
2-Tier Transact Pattern



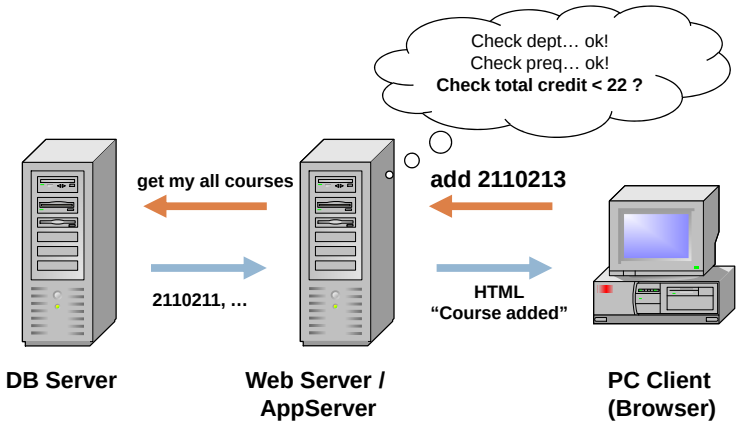
- Pros:
  - Simple to develop (initially).
- Cons:
  - Poor scalability.
  - Poor manageability and deployability.

3/N-Tier Transact Pattern

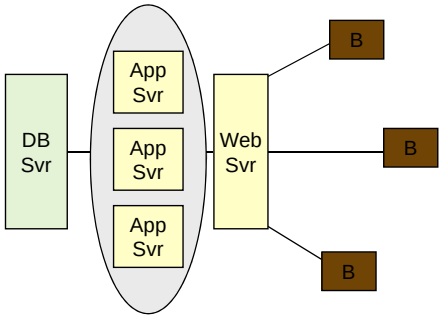
- Database + Web server / App server + Browsers (Thin clients).



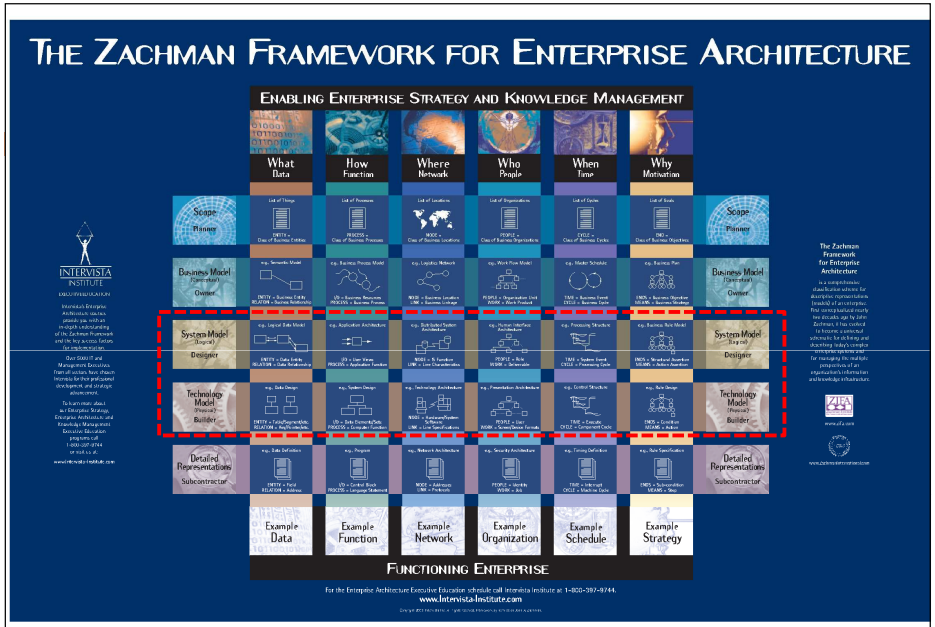
3/N-Tier Registration System



3/N-Tier Transact Pattern



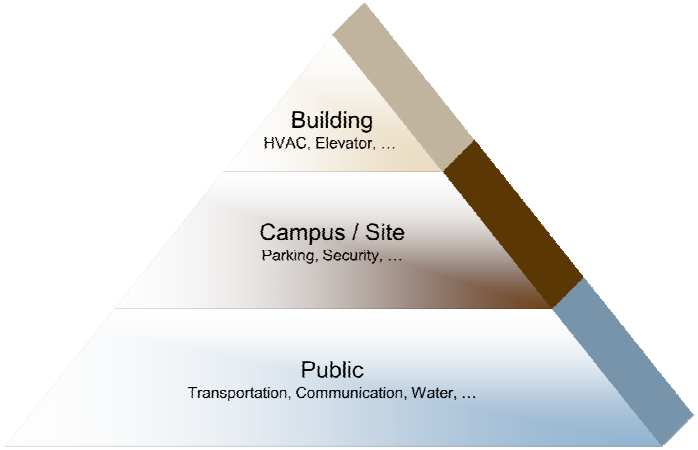
- Pros:
  - Scalability.
  - Flexible.
  - Simple to manage.
  - Great for integration.
- Cons:
  - Slightly difficult to develop (initially).



What is Infrastructure ?

- Infrastructure:
  "the structure beneath a structure"
  - public utilities e.g. water, electricity, telephone, etc.
  - Infrastructure is usually shared and layer-based.
  - Lower layers are more static and permanent than upper layers.
  - Layers are independent in term of lifecycles (plan, build, run, change, exit) and ownership.

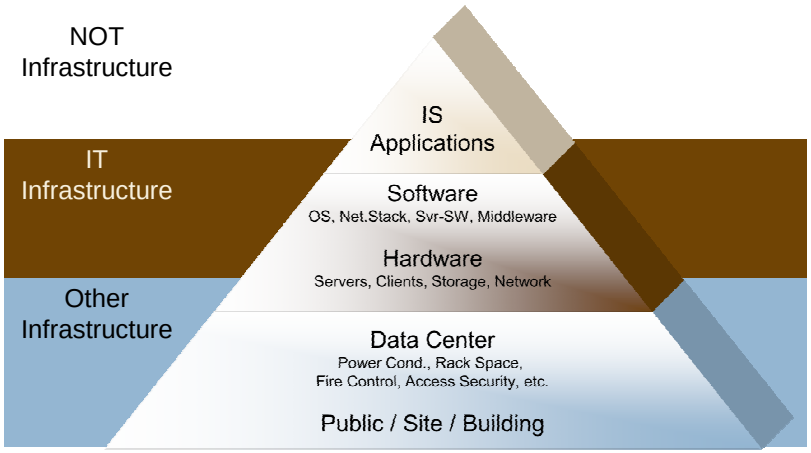
Layers of Infrastructure



What is IT Infrastructure ?

- IT Infrastructure:
  "the structure of IT systems that supports IS, which comprise of IT components, the externally visible properties of those components, and the relationships among them"
  - Consist of both Hardware and Software.
  - Always rely on other infrastructure e.g. Data Center, Building, etc.

Layers of IT Infrastructure



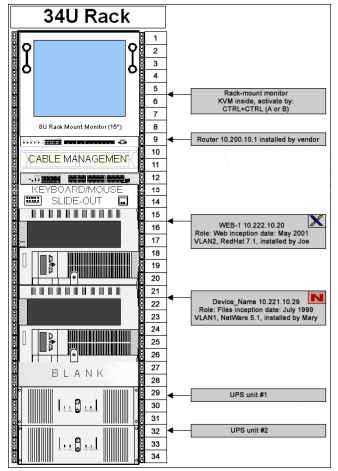
Base Infrastructure

- **Servers:**
  - Web server, Application server, Database server, …
- **Storage:**
  - Main storage, secondary storage, backup, …
- **Networks:**
  - Internet, Intranet, proprietary networks, …
- **Security:**
  - Network security equipment, data encryption, …

Server

- Workhorse of IT system
  - shares a resource (e.g. files) to one or more clients
- Must be high-performance, robust, and reliable
  - Using high-grade components
  - Fast and large RAM
  - High I/O (e.g. FC) and network bandwidths
  - Redundant power supply
- Often installed on racks

Rack-Mount Server



Mainframe: The Big Iron

- Large, powerful computers for running many different tasks at the same time
- Highly reliable and secured servers
- Extensive input-output facilities
- Backward compatibility with older software
- Support massive throughput

Real World: IBM System Z10

- Announced in 2008
- Capacity
  - Up to 64 x Quad-core (4.4 GHz) CPUs
  - Up to 1.5 TB main memory

Mainframe vs. Supercomputer

Mainframe

- optimized for simple computations with huge amounts of external data (e.g. payroll processing)
- Can handle a wider variety of tasks
- Good at batch processing, such as billing

Supercomputer

- optimized for complicated computations that take place largely in memory (e.g. weather forecasting)
- built for one or a very few specific institutional tasks (e.g. simulation and modeling)

Blade Server

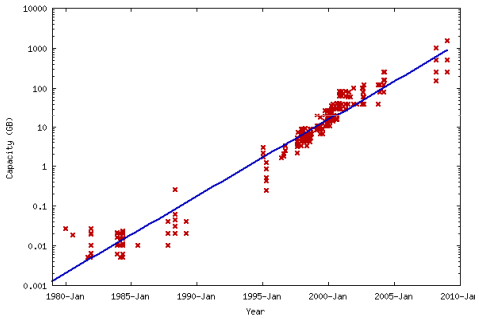
- Stripped down computer servers with a modular design
- Hosted in blade enclosure
  - Support multiple blade servers
  - Provides power, cooling, networking, interconnects, and management
- Benefits
  - Space and power consumption optimization

Blade Server

Blade Server

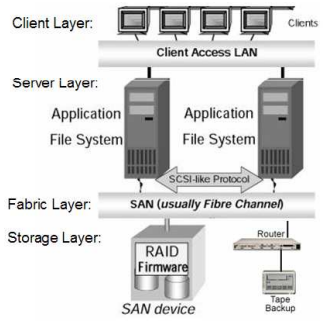
Storage (Hard Disk Drive)

- Capacity
- Interface Technology
  - IDE (P-ATA) / SATA
  - SCSI / SAS
  - FC
- Performance
  - Seek Time
  - Data Transfer Rate (e.g. 15K RPM)

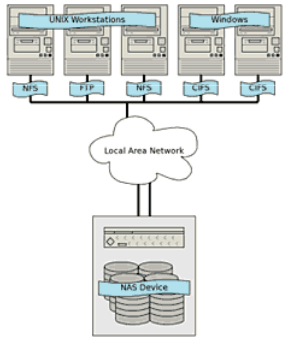


Storage Architecture: SAN vs. NAS

SAN: Storage Area Network



NAS: Network Attached Storage

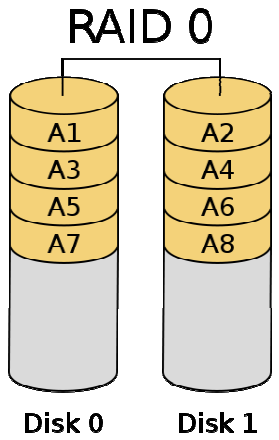


RAID

- Redundant Array of Inexpensive Disks
  - Achieve high levels of storage reliability
  - Using low-cost and less reliable PC-class disk-drive components
- Hot-Spare
  - A drive physically installed in the array which is inactive until an active drive fails
- Hot-Swapped
  - Ability to add/remove disks without shutting down the system

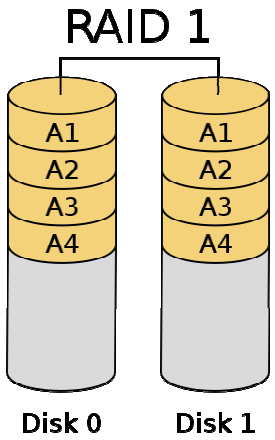
Popular RAID Level

- RAID 0 – Striping
  - improved performance
  - additional storage
  - no redundancy or fault tolerance
  - N storage capacity



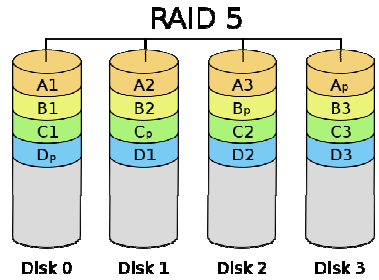
Popular RAID Level

- RAID 1 – Mirroring
  - Provides fault tolerance from disk errors
  - Up to one-disk failure
  - Increased read performance
  - Very small performance reduction when writing
  - 1 storage capacity



Popular RAID Level

- RAID 5 – Interleave Parity
  - Distributed data to all disks with one disk as a parity container
  - Good disk performance
  - Up to one-disk failure
  - N-1 storage capacity



Real World: Google Infrastructure

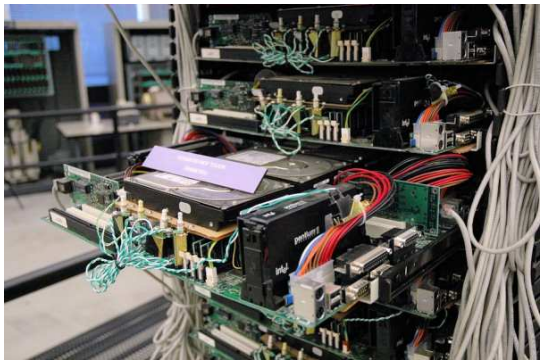
- Workload
  - Hundreds of millions of search request per day
  - Process about 1 petabyte of user-generated data every hour
  - Support various Google products
    - Search, advertising, email, maps, video, chat, blogger

Google™

Real World: Google Infrastructure

- Develop its own distributed systems infrastructure
  - GFS
  - MapReduce
  - BigTable
- Use commodity hardware for cost-effective solution
  - Linux, in-house rack design, PC class mother boards, low end storage
  - Estimated 450,000 low-cost commodity servers in 2006

Real World: Google Infrastructure



Real World: Google Infrastructure

Data Center Map (as of April 2008)

- spent on data centers
  - $1.9 billion in 2006
  - $2.4 billion in 2007
- cost on the order of US$2 million per month in electricity charges

Location Selection Criteria

- Large volumes of cheap electricity
- Green energy / renewable power sources
- Proximity to rivers and lakes for cooling purposes
- Large areas of land for more privacy and security
- Distance to other Google data centers (for fast connections)
- Tax incentives

Issues of ISA

- Infrastructure choices
  - Transact patterns
  - Centralized or distributed
  - Programming languages and development
- Security
  - Internet is a global network
  - Security becomes very important
  - Security infrastructure
    - Attacking prevention: firewall, VPN, IDS
    - Secured data transfer: encryption algorithms, security protocols

Issues (2)

- Performance
  - Responsiveness
    - How long (turn-around time) can a user tolerate ?
  - Scalability
    - If we have more users, can the system still respond within a certain period of time ?
    - If we have new services, can the system provide new services while it still meets the responsiveness criteria.

Issues (3)

- Manageability
  - Operational
    - Reliability / Availability / Downtime (including maintenance period)
    - Cooling issues
  - Cost
    - Electricity
  - Maintainability
    - How difficult is it for the admins to deploy, maintain, and upgrade the system ?

Summary

- Transact Pattern: Tier-Based Architecture
- IT Infrastructure
  - Server
  - Storage
- Real World: Google Infrastructure
- Infrastructure Issues

References

- T. Hoff, "Google Architecture", http://highscalability.com/google-architecture, 23 Nov 2008.
- "Google Platform", http://en.wikipedia.org/wiki/Google_platform
- E. Turban *et al.*, *Information Technology for Management: Transforming Business in the Digital Economy*, John Wiley & Sons, Inc., 2002.
- B. Robertson and V. Sribar, *The Adaptive Enterprise*, Intel Press, January 2002.
- R. Peacock, "Distributed Architecture Technologies", *IEEE IT Professional*, May/June 2000, pp. 58-60.