IS ARCHITECTURE OVERVIEW

2110684 Information System Architecture

Natawut Nupairoj, Ph.D.

Course Plan

- Overview and Background Concepts
- Core Infrastructure
- System Management / Security / Tuning

Things That Architects Can Do

- Plan technology direction and set technology standards
  - Help you figure out which technologies you should support.
- Review plans, designs and purchases
  - Assess how well a plan aligns with current direction and desired future positions.
- Identify opportunities to reuse components and services.
  - Leverage enterprise contracts and license agreements.
  - Integrate shared services where they might be cost-effective.
- Review business organization and business processes
  - Technical Architecture: align your technology plan with enterprise goals, business plans and business processes.
  - Enterprise Architecture: align your business plans, business process and technology plan with your enterprise goals.

From Don Jerman, “Architecture Review Processes”

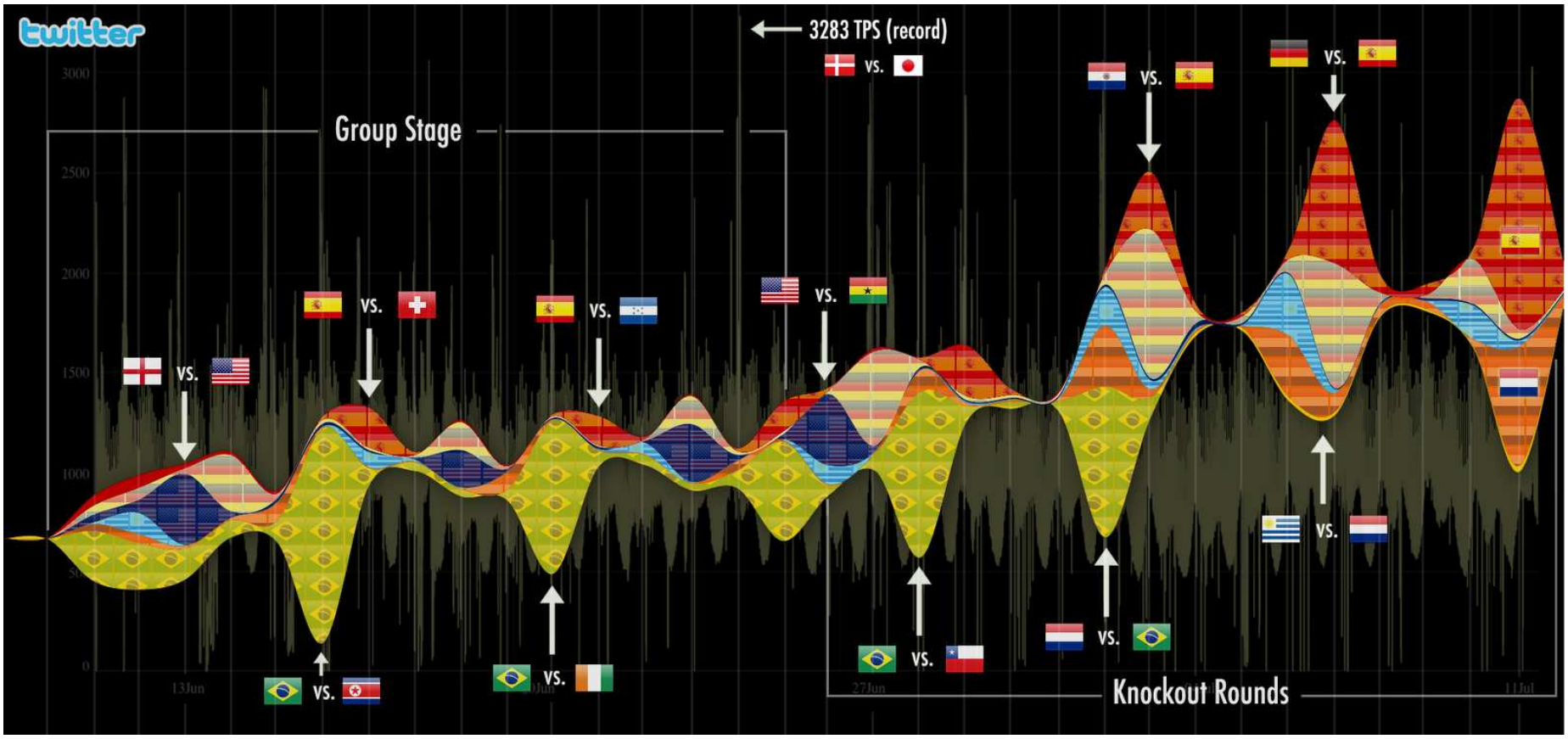
Most common information today?

- 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.

Real World Example: Twitter

- SMS of the Internet
- Sending short 140-character message to followers
- Start in 2006
- From 120,000 tweets/month (in 2007) to 1,500,000,000 tweets/day (in 2010) - 750 tweets/second
- 300,000 new subscribers a day
- Just reach 20,000,000,000 tweets in July 31st, 2010 (a Japanese graphic designer)

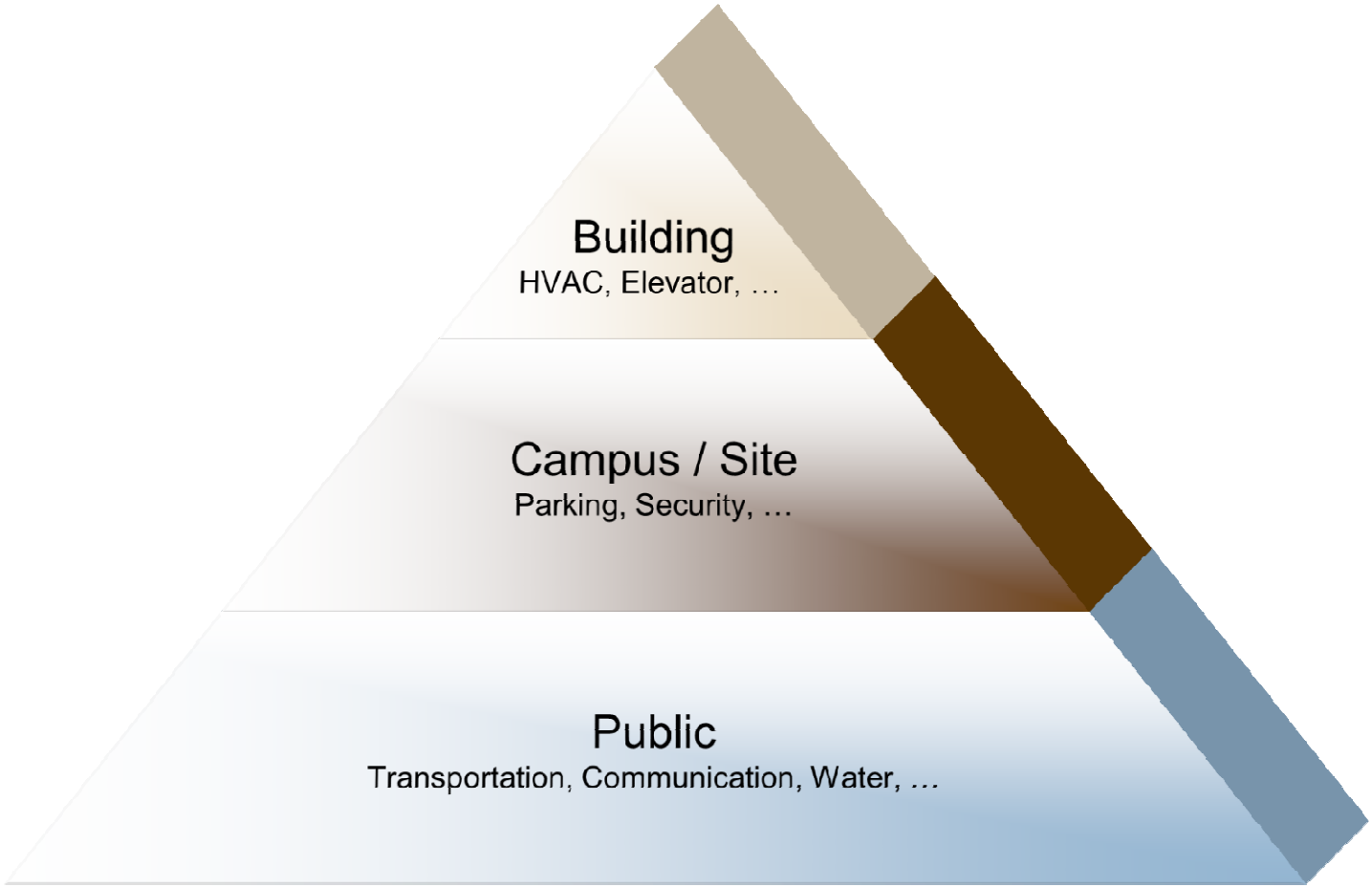
Twitter vs. World Cup 2010



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.

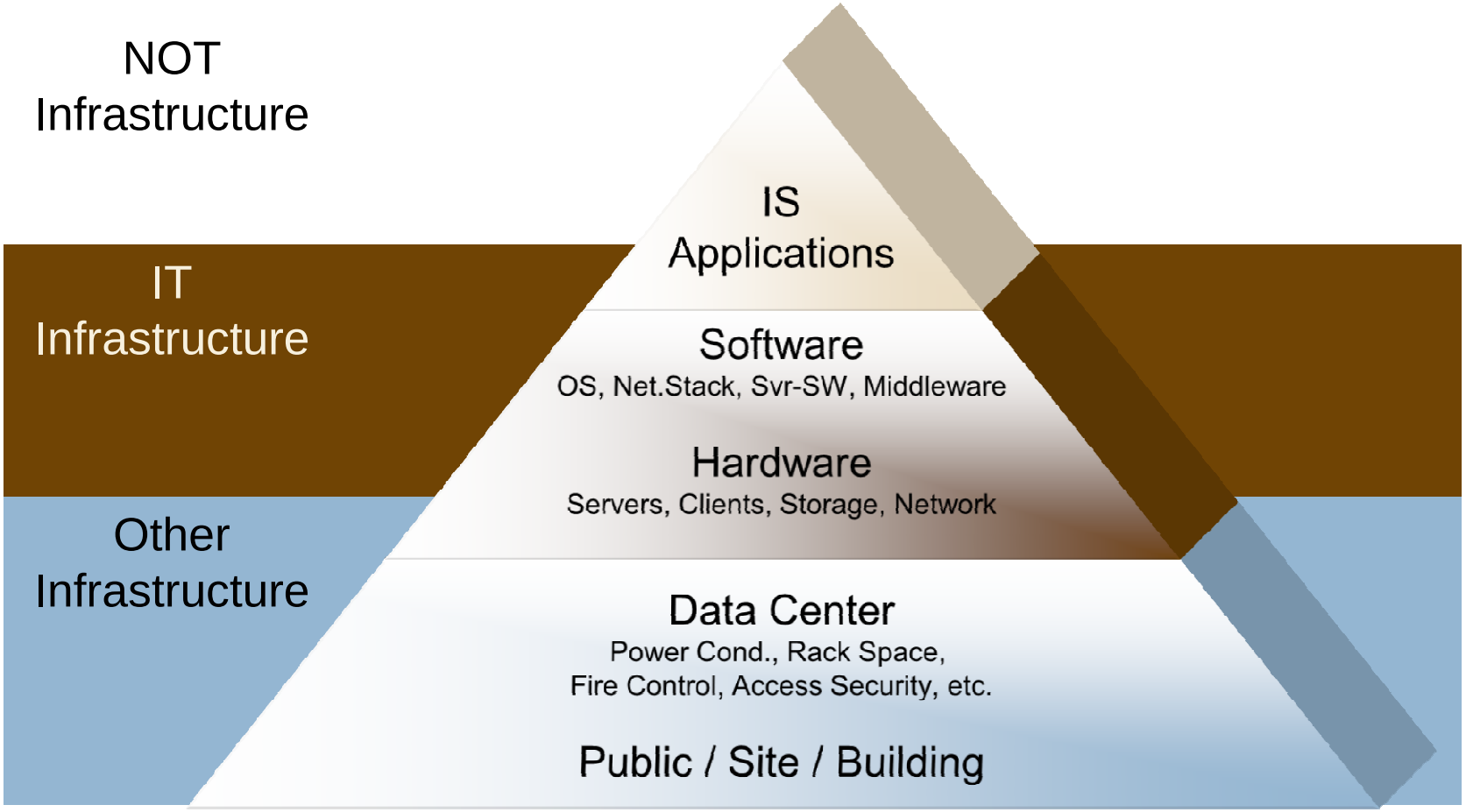
Layers of Infrastructure



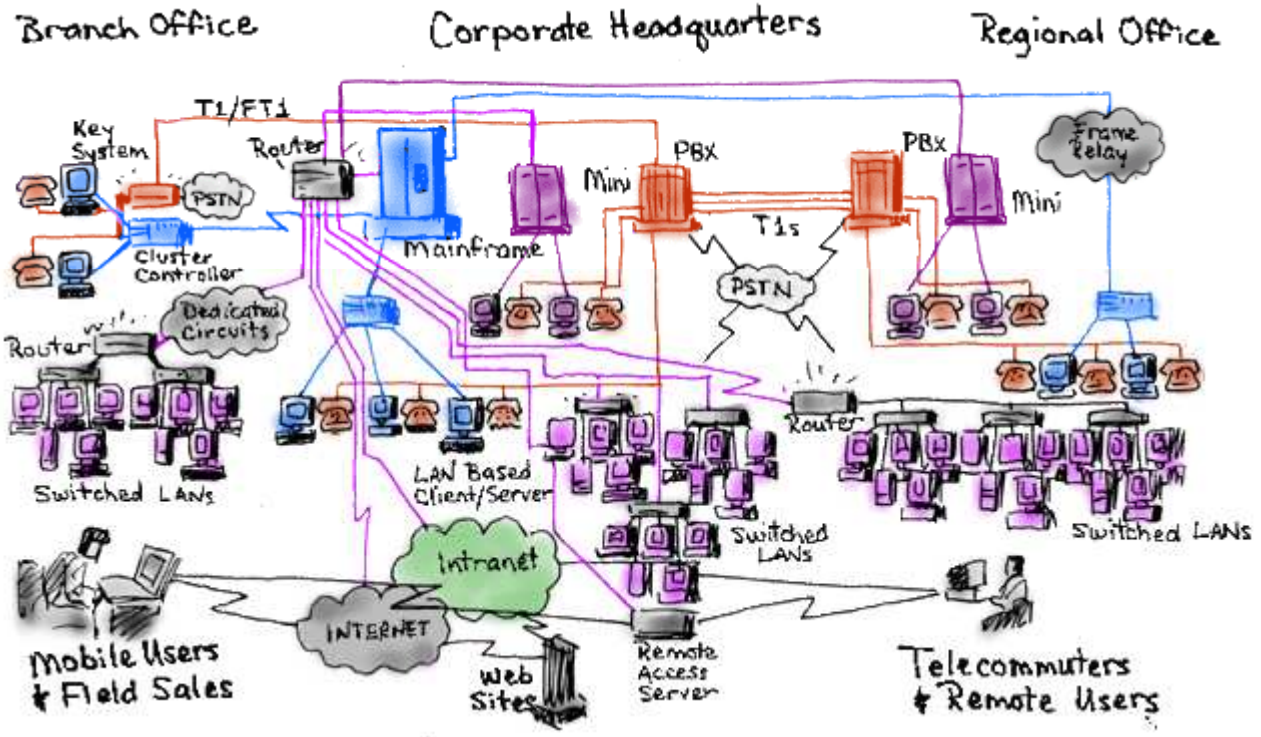
What is IT Infrastructure ?

- 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



IT Infrastructure



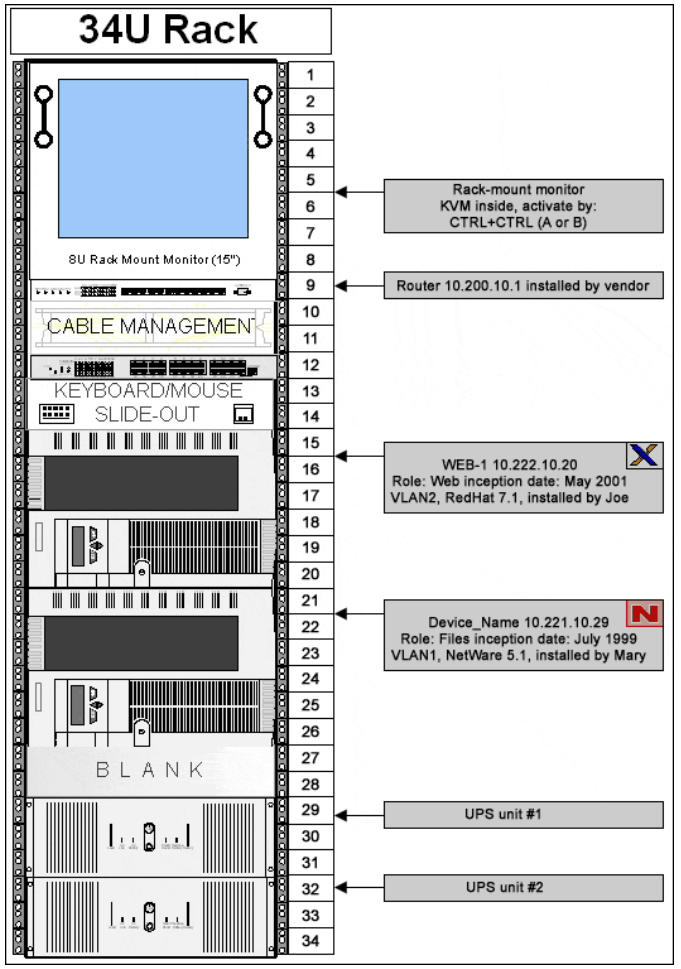
Base Infrastructure

- **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

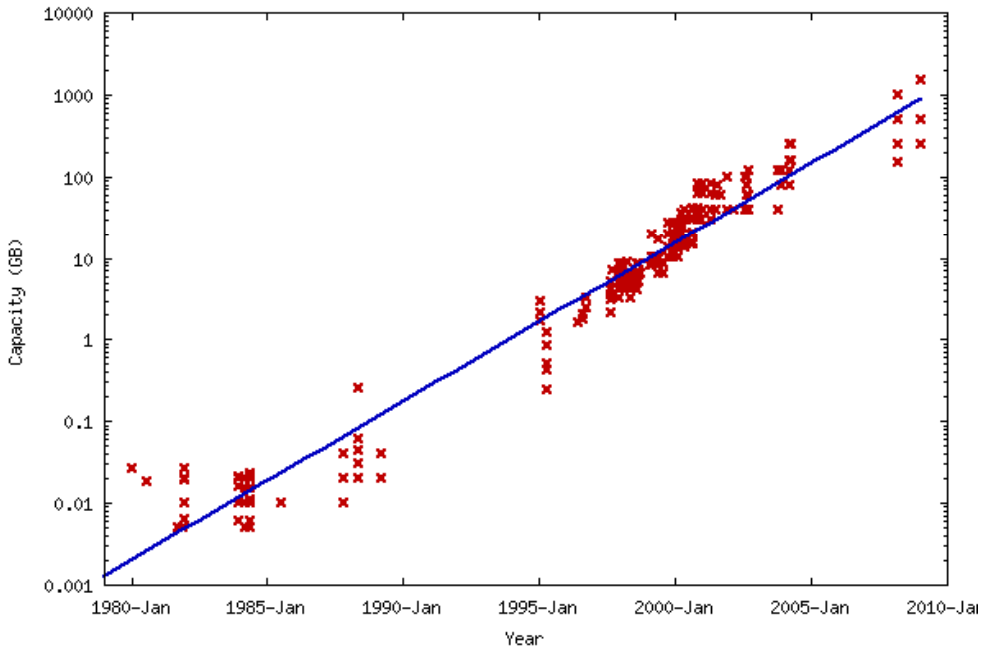
- Stripped down computer servers with a modular design
- Hosted in blade enclosure
  - Support multiple blade servers
  - Provides power, cooling, networking, interconnects, and management
- Benefits
  - Space and power consumption optimization

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)



Network Attached Storage (NAS)

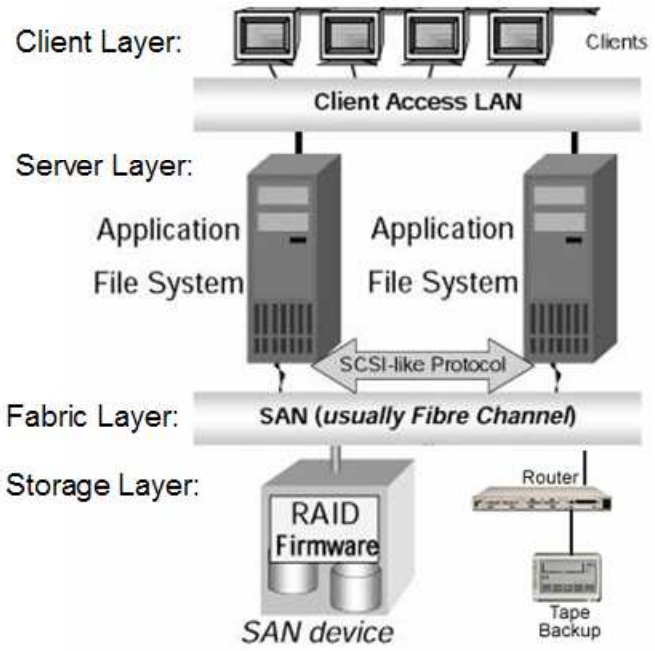
- File-level data storage
- Connecting directly to standard network
- Standard file-based protocols
  - NFS, CIFS, FTP, HTTP
  - UPnP, Rsync, …
- Pros: Simple to operate and maintain, Cheap
- Cons: Performance limitation

Storage Area Network (SAN)

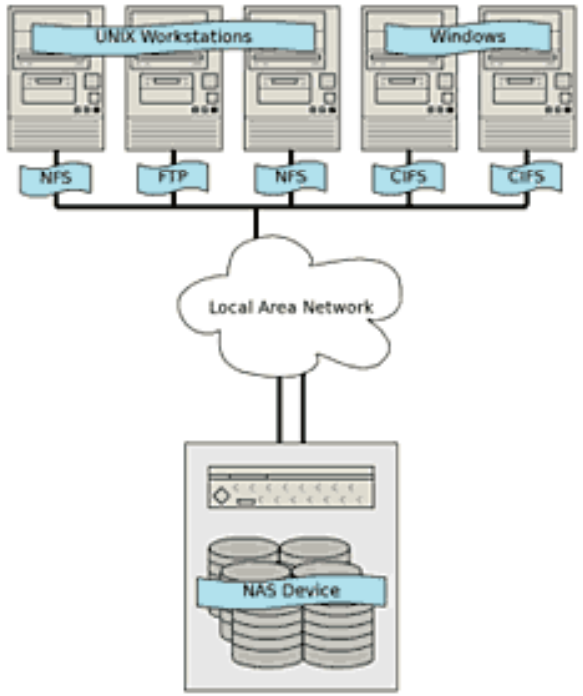
- Block-level data storage
- Connect to proprietary "SAN" network
- Storage protocols
  - SCSI
  - Fiber Channel
  - FICON
- Pros: High performance
- Cons: Expensive, complex

Storage Architecture: SAN vs. NAS

SAN: Storage Area Network



NAS: Network Attached Storage

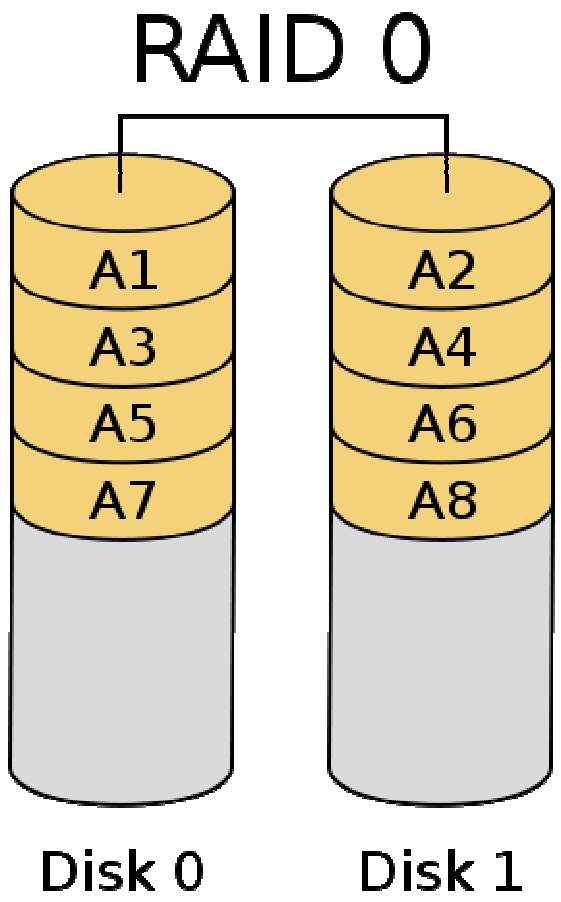


RAID

- 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

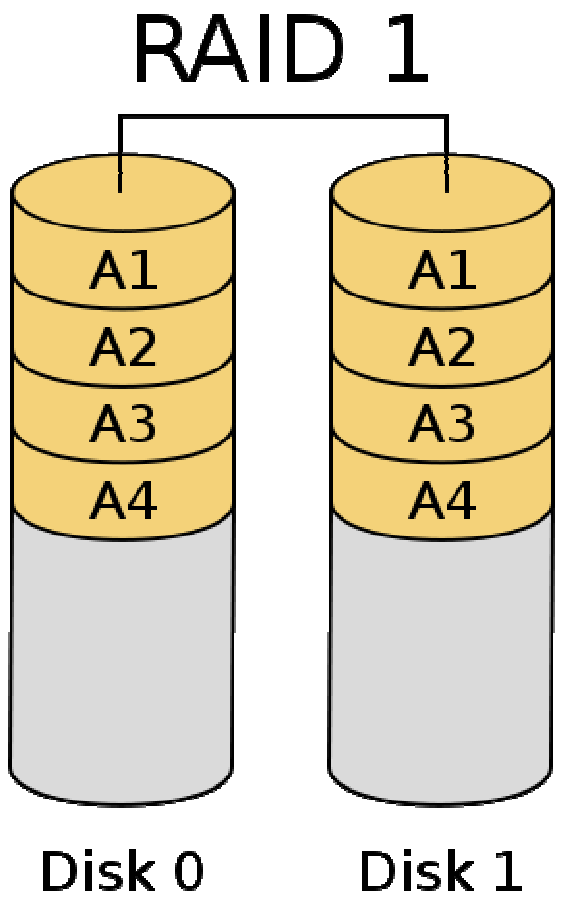
Popular RAID Level

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



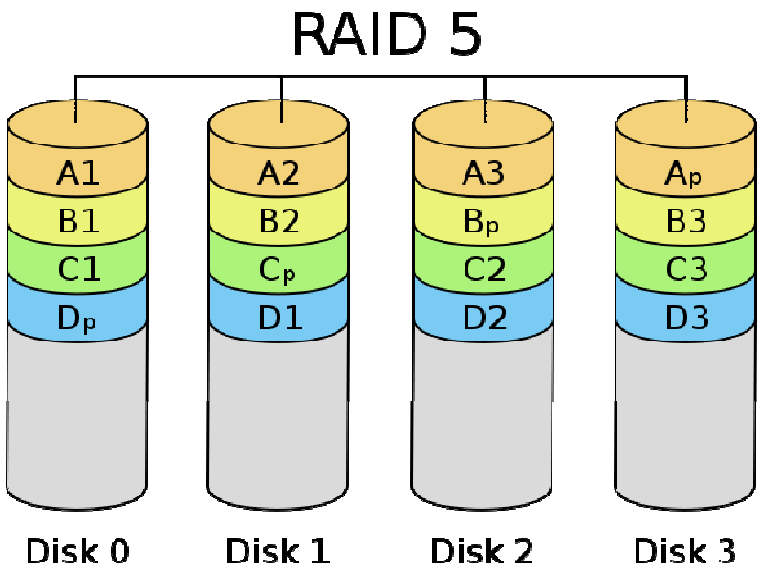
Popular RAID Level

- 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



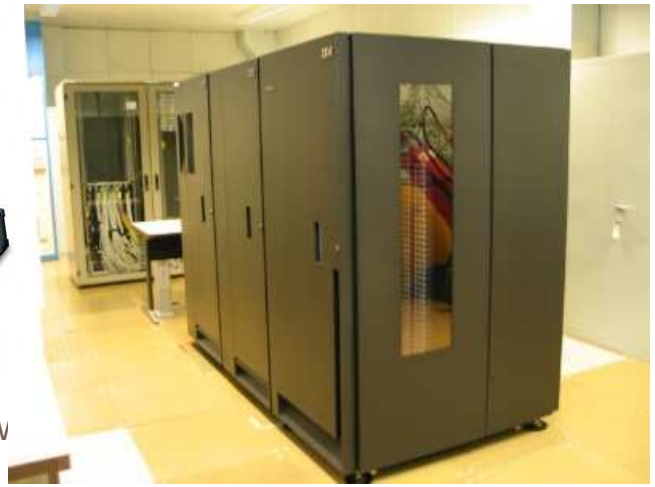
Popular RAID Level

- 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



Backup

- Make copies of data to be restored when disaster occurs or files are accidentally deleted
- Wide-range of medias
  - Magnetic tapes
  - Hard disks
  - Remote backup services
- Full vs. Incremental

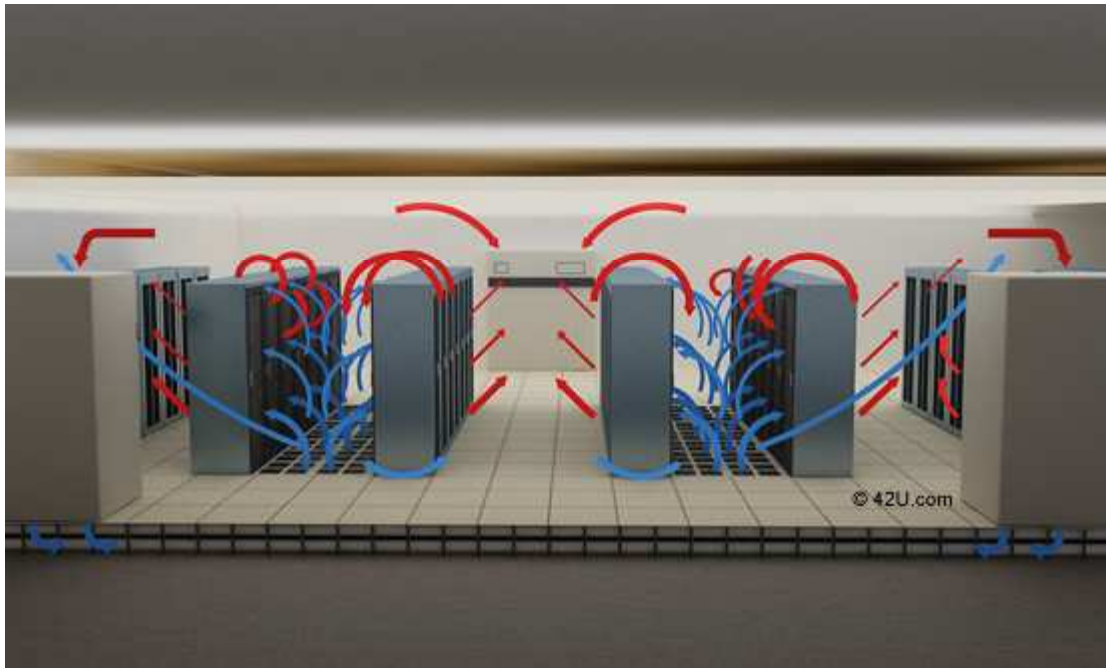


Data Center

- A room / floor / building that hosts computer systems
- Provide
  - Power supplies
  - Network
  - Environmental controls
  - Security

Data Center Components

Data Center Components



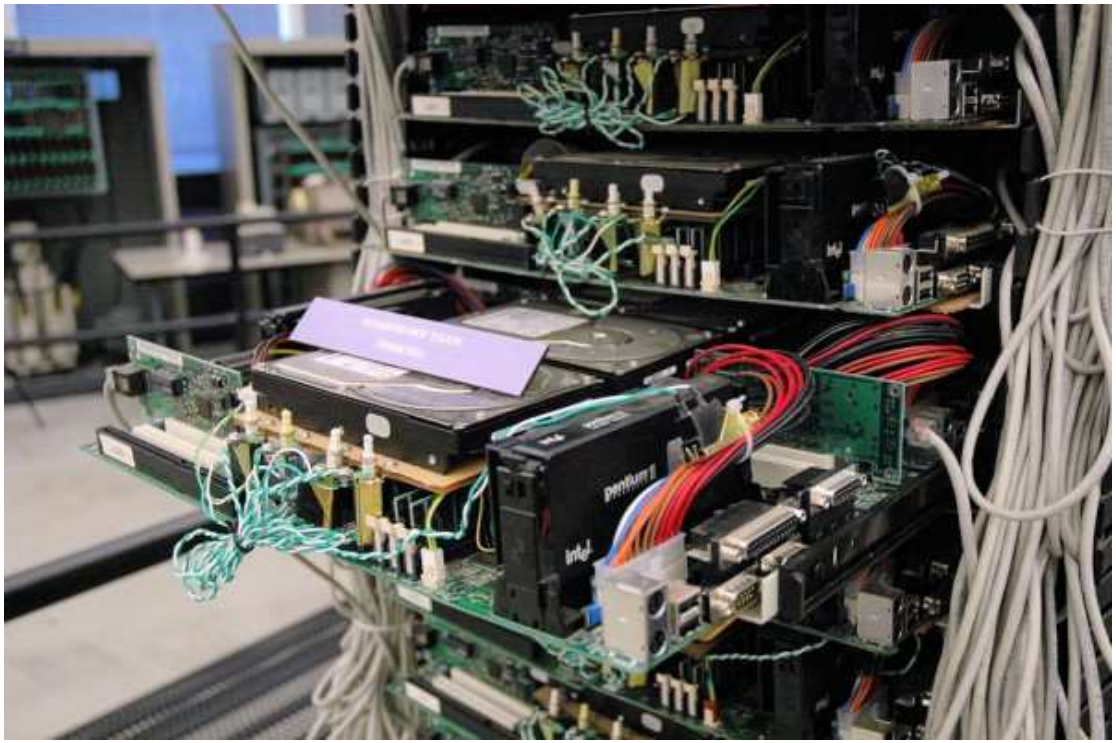
Real World: Google Infrastructure

- 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

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.