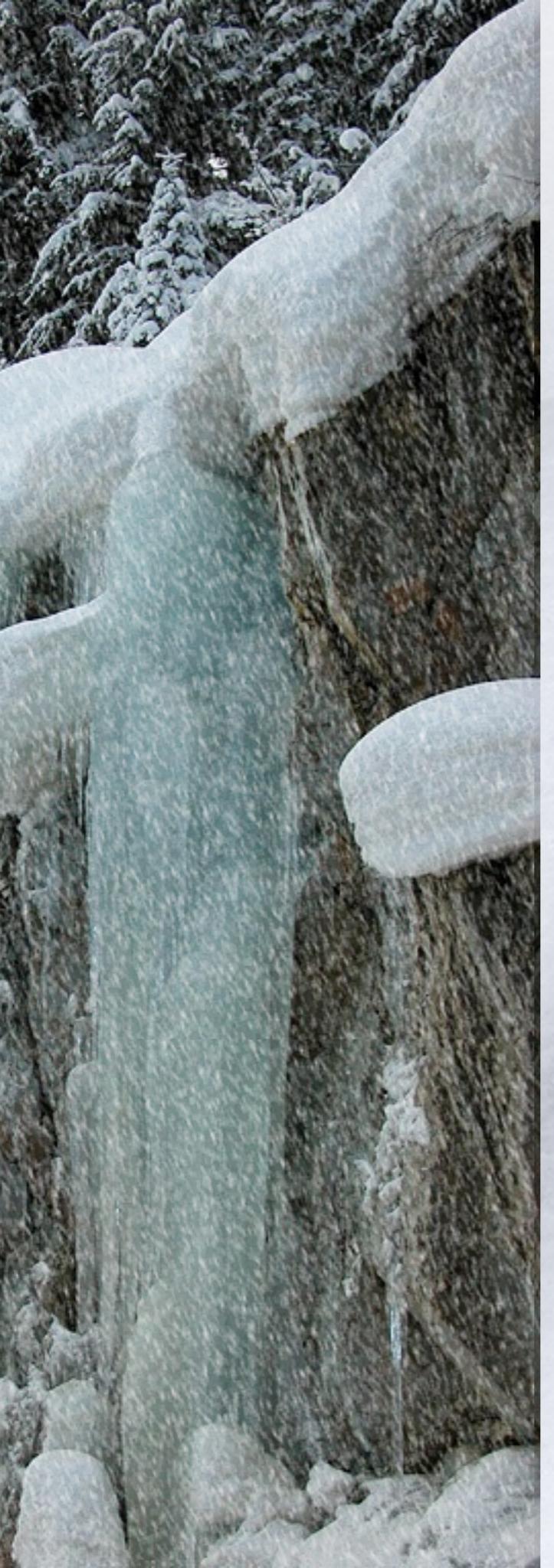


Lecture 1

Internet and World Wide Web

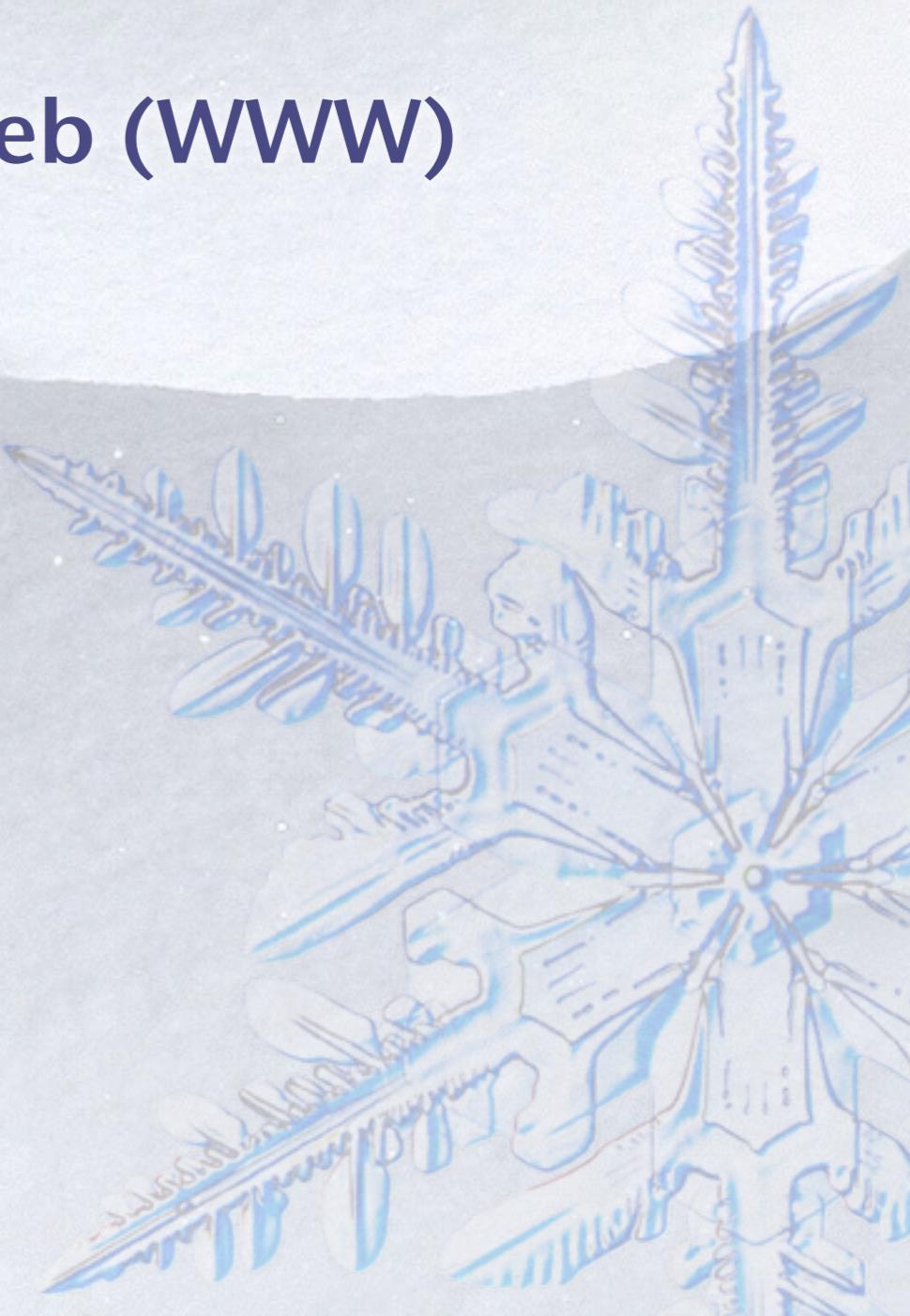




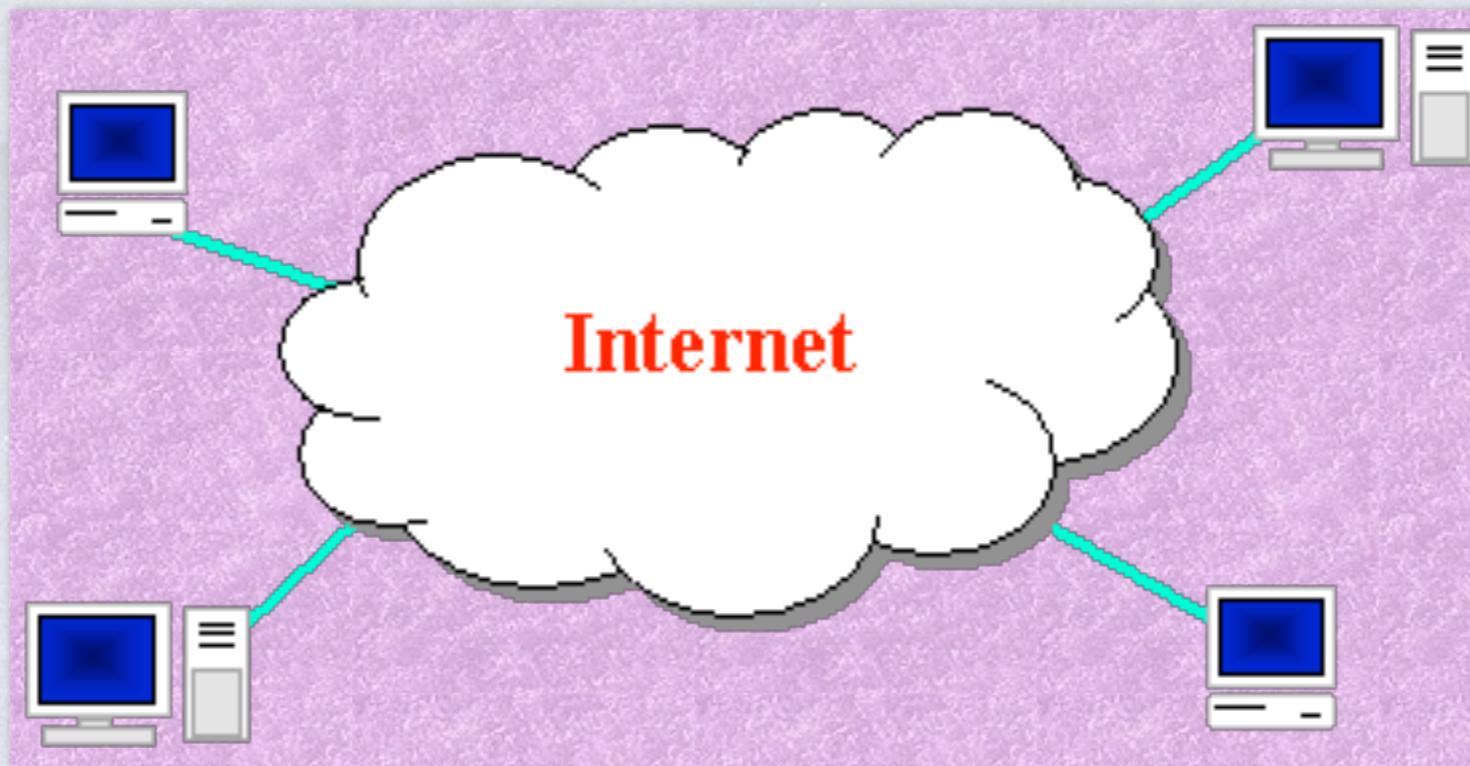
Outline

* **The Internet**

* **The World Wide Web (WWW)**



The Internet



* Wikipedia: <http://en.wikipedia.org/wiki/Internet>

- * a connection of computer networks using the Internet Protocol Suite (TCP/IP)

* What's the difference between the Internet and the World Wide Web (WWW)?

- * WWW = HTML* + HTTP(S)
- * including CSS, JavaScript, and other browser enabled content

Brief History

- * began as a US Department of Defense network called ARPANET (1960s-70s)
- * initial services: electronic mail, file transfer
- * opened to commercial interests in late 1980s
- * WWW created in 1989-91 by Tim Berners-Lee
 - * This is for Everyone
- * popular web browsers released:
 - * Netscape 1994, IE 1995
- * Amazon.com opens in 1995; Google January 1996
- * Chinese First Connection with Internet: Chinese Academics Net, by Computer Applying Technology Institute of Beijing 1986,
- * First email, Sep. 14 1987, from CATIB, “Across the Great Wall we can reach every corner in the world”
- * Chinese First Full Internet Connection: NCFC (National Computing and Networking Facility of China) 1994
- * Baidu 1999, Taobao 2003



Key aspects of the Internet

- ❖ Internet is for freedom of information
- ❖ computers can dynamically join and leave the network
- ❖ built on open standards; anyone can create a new device
- ❖ lack of centralized control (mostly)
- ❖ everyone can use it with simple, commonly available software
- ❖ internet Vs. Internet

People and Organizations

❖ Internet Engineering Task Force (IETF)

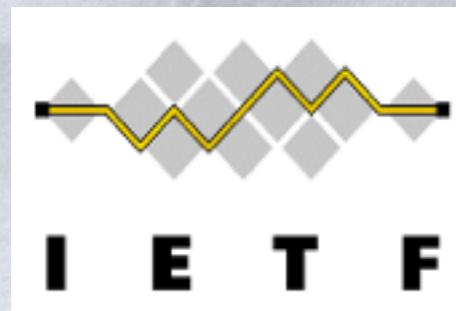
- * internet protocol standards

❖ Internet Corporation for Assigned Names and Numbers (ICANN)

- * decides top-level domain names

❖ World Wide Web Consortium (W3C)

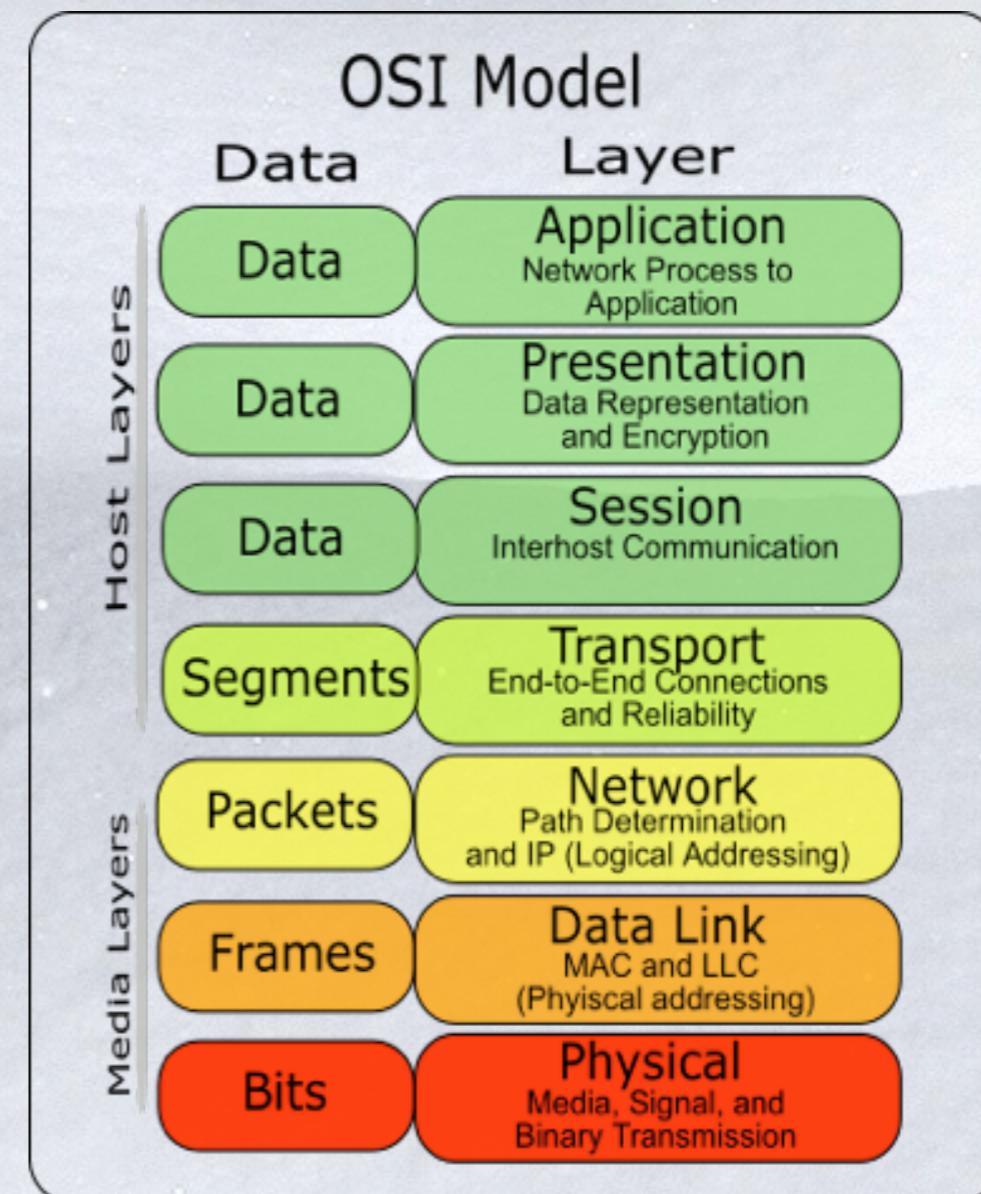
- * Web standards



Layered architecture

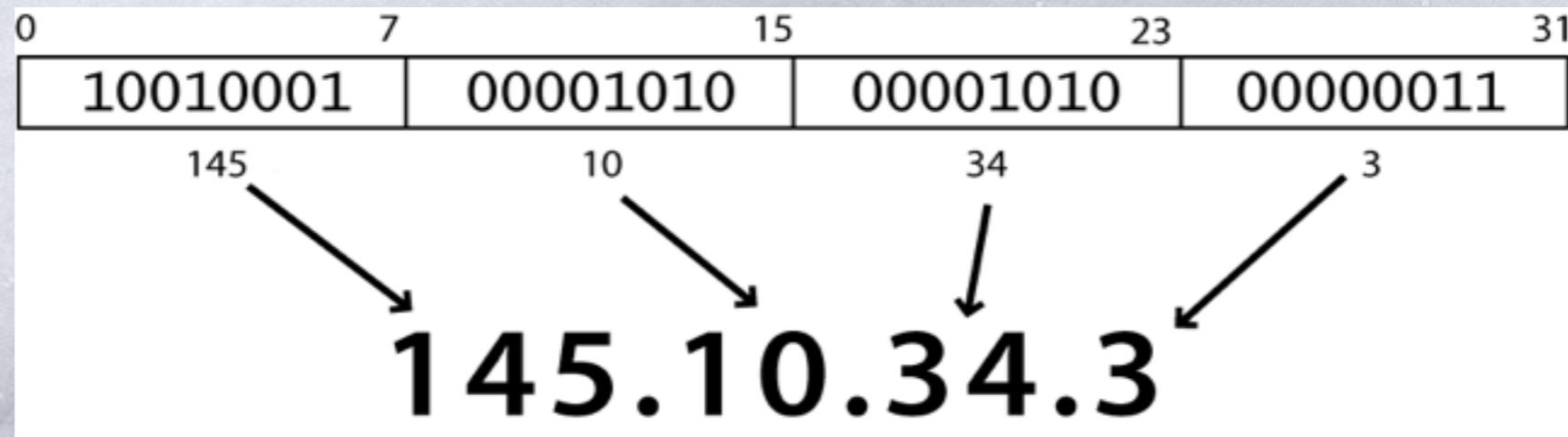
The internet uses a layered hardware/software architecture (OSI model):

- * physical layer: devices such as coaxial cables, fiber-optic lines, modems
- * data link layer: basic hardware protocols (Ethernet, Wi-Fi, DSL, ATM, PPP)
- * network / internet layer: basic software protocol (IP)
- * transport layer: add reliability to network layer (TCP, UDP)
- * application layer: implements specific communication for each kind of program (HTTP, POP3/IMAP, SSH, FTP)



Internet Protocol (IP)

- ❖ the IP is the underlying system of communication for all data (packets) sent across the internet.



Classful Addressing

- ❖ When IP was first standardized in Sep 1981, each system attached to the IP based Internet had to be assigned a unique 32-bit address
- ❖ The 32-bit IP addressing scheme involves a two level addressing hierarchy

Network Number/Prefix

Host Number

Classful Addressing...

- ❖ Divided into 5 classes
- ❖ Class A 8 bits N/W id and 24 bits host id and so on B,C.
- ❖ Wastage of IP addresses by assigning blocks of addresses which fall along octet boundaries

	First byte	Second byte	Third byte	Fourth byte
Class A	0			
Class B	10			
Class C	110			
Class D	1110			
Class E	1111			

Techniques to reduce address shortage in IPv4

- ❖ Subnetting
- ❖ Classless Inter Domain Routing (CIDR)
- ❖ Network Address Translation (NAT)

Subnetting

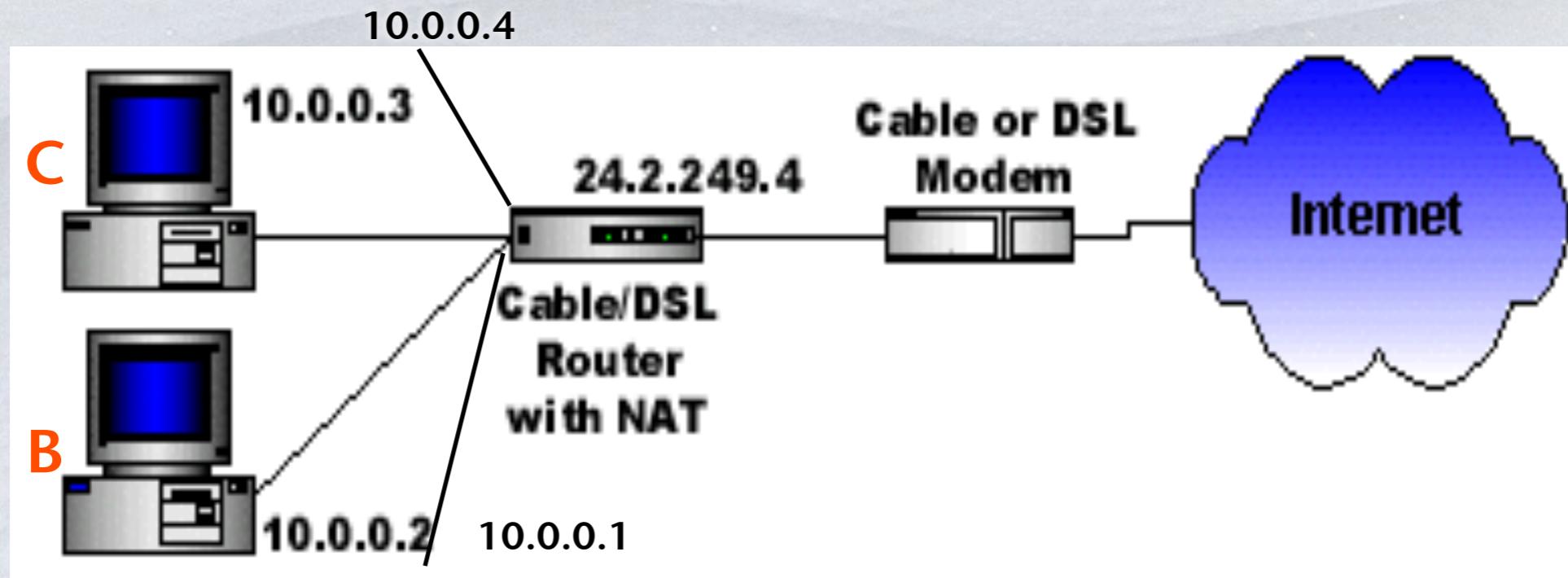
- ❖ Three-level hierarchy: network, subnet, and host.
- ❖ The extended-network-prefix is composed of the classful network-prefix and the subnet-number
- ❖ The extended-network-prefix has traditionally been identified by the subnet mask

Network-Prefix	Subnet-Number	Host-Number
----------------	---------------	-------------

Classless Inter-Domain Routing

- ❖ Eliminates traditional classful IP routing.
- ❖ Supports the deployment of arbitrarily sized networks
- ❖ Routing information is advertised with a bit mask/ prefix length
 - * specifies the number of leftmost contiguous bits in the network portion of each routing table entry
- ❖ Example: 192.168.0.0/21

NAT Example



Source Computer	Source Computer's IP Address	Source Computer's Port	NAT Router's IP Address	NAT Router's Assigned Port Number
A	10.0.0.1	400	24.2.249.4	1
B	10.0.0.2	50	24.2.249.4	2
C	10.0.0.3	3750	24.2.249.4	3
D	10.0.0.4	206	24.2.249.4	4

Features of IPv6

- ❖ Larger Address Space
- ❖ Aggregation-based address hierarchy
 - * Efficient backbone routing
- ❖ Efficient and Extensible IP datagram
- ❖ Stateless Address Autoconfiguration
- ❖ Security (IPsec mandatory)
- ❖ Mobility

128-bit IPv6 Address

3FFE:085B:1F1F:0000:0000:0000:**00A9:1234**

8 groups of 16-bit hexadecimal numbers separated by “::”

Leading zeros can be removed

3FFE:85B:1F1F::A9:1234

:: = all zeros in one or more group of 16-bit hexadecimal numbers

Transmission Control Protocol (TCP)

- ❖ adds multiple, guaranteed message delivery on top of IP
- ❖ multiplexing: multiple programs using the same IP address
 - * port: a number given to each program or service
 - * 80: Web browser (443 for secure browsing)
 - * 25: email
 - * 22: ssh
 - * 21: ftp
 - * more common ports...
- ❖ some programs (QQ, games, streaming media programs) use simpler UDP protocol instead of TCP
- ❖ find out ports used:
 - * in a terminal, using netstat (Windows) command
 - * using CurrPorts

The Domain Name System

- ❖ The domain name system is usually used to translate a host name into an IP address .
- ❖ Domain names comprise a hierarchy so that names are unique, yet easy to remember.

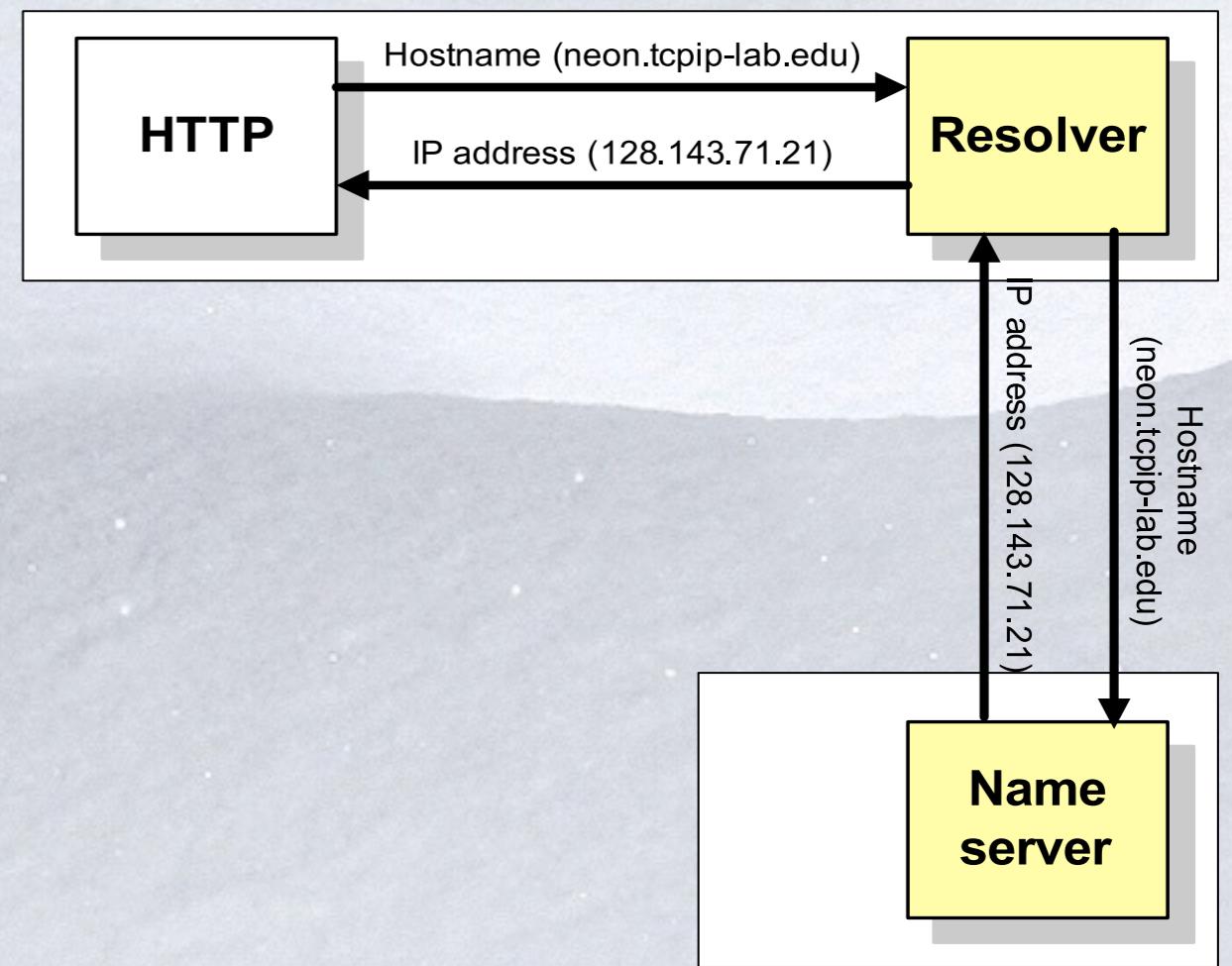
Before there was DNS

.... there was the HOSTS.TXT file

- ✿ Before DNS (until 1985), the name-to-IP address was done by downloading a single file (hosts.txt) from a central server with FTP.
 - * Names in hosts.txt are not structured.
 - * The hosts.txt file still works on most operating systems. It can be used to define local names.
 - * Windows: C:\Windows\system32\drivers\etc\hosts
 - * Mac: /private/etc/hosts
 - * Linux: /etc/hosts

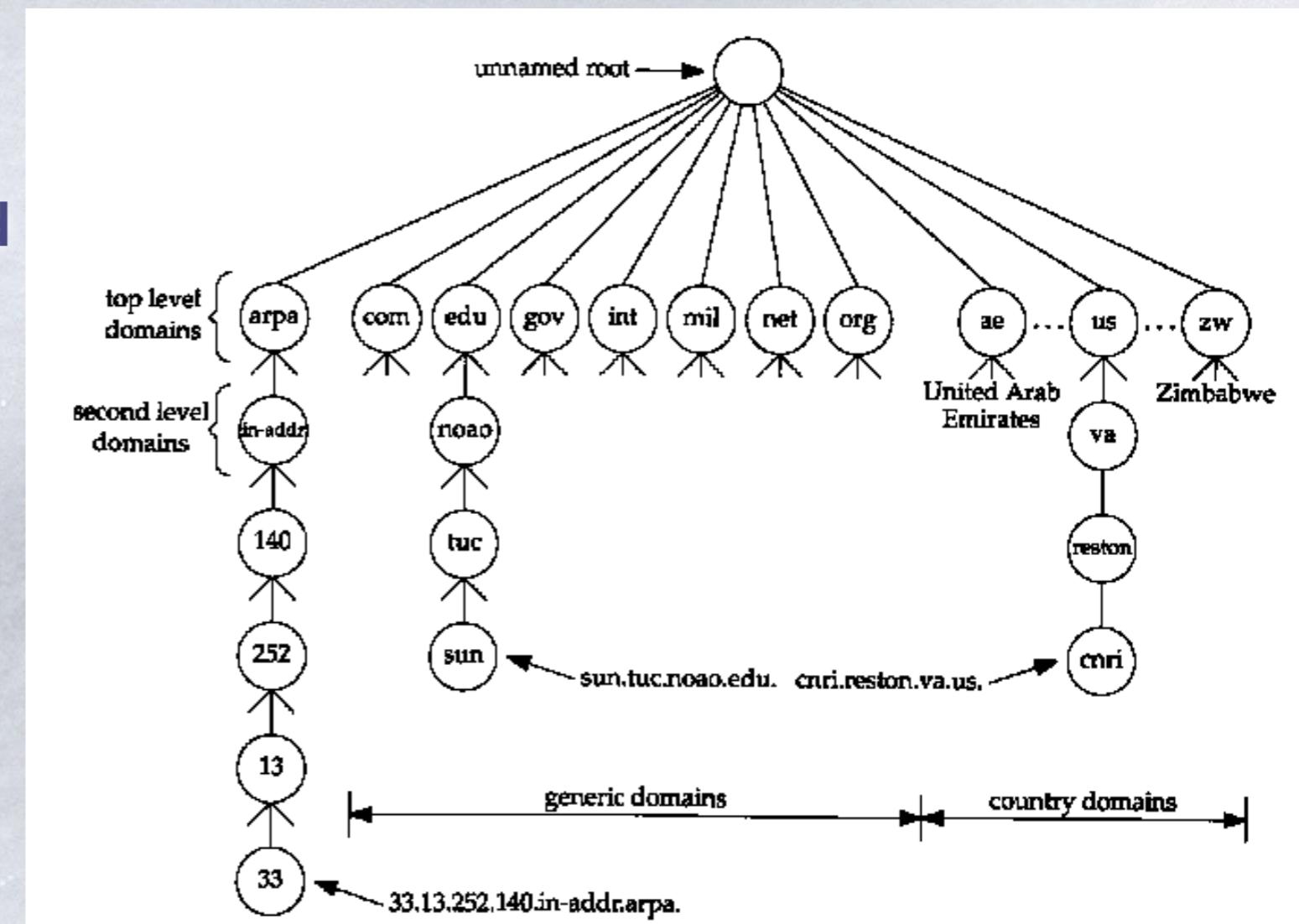
Resolver and Name Server

- ❖ An application program on a host accesses the domain system through a DNS client, called the resolver
- ❖ Resolver contacts DNS server, called name server
- ❖ DNS server returns IP address to resolver which passes the IP address to application
- ❖ Reverse lookups are also possible, i.e., find the hostname given an IP address



DNS Name Hierarchy

- ❖ DNS hierarchy can be represented by a tree
- ❖ Root and top-level domains are administered by an Internet central name registration authority (ICANN)
- ❖ Below top-level domain, administration of name space is delegated to organizations
- ❖ Each organization can delegate further



Domain names

- ❖ Hosts and DNS domains are named based on their position in the domain tree
- ❖ Every node in the DNS domain tree can be identified by a unique Fully Qualified Domain Name (FQDN). The FQDN gives the position in the DNS tree.

software.nju.edu.cn or software.nju.edu.cn.

- ❖ A FQDN consists of labels (“software”, “nju”, “edu”, “cn”) separated by a period (“.”)
 - * There can be a period (“.”) at the end.
 - * Each label can be up to 63 characters long
 - * FQDN contains characters, numerals, and dash character (“-”)
 - * FQDNs are not case-sensitive

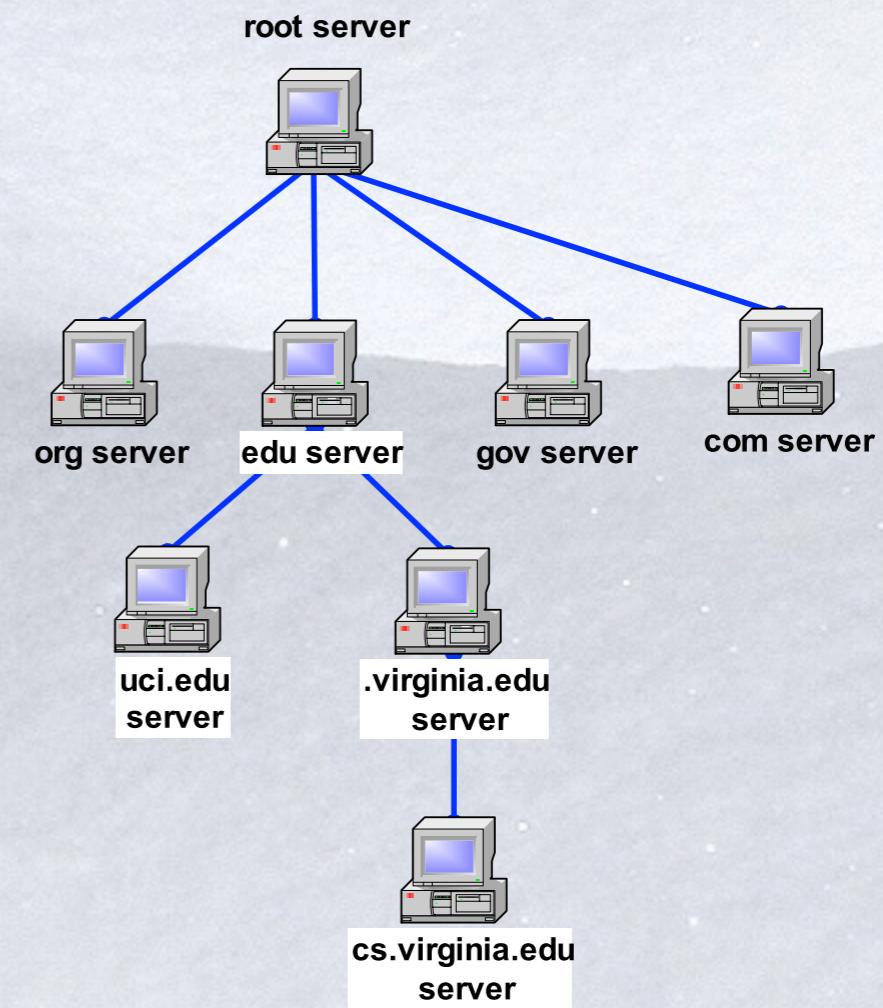
Hierarchy of name servers

- ❖ The resolution of the hierarchical name space is done by a hierarchy of name servers

- ❖ Each server is responsible (authoritative) for a contiguous portion of the DNS namespace, called a zone.

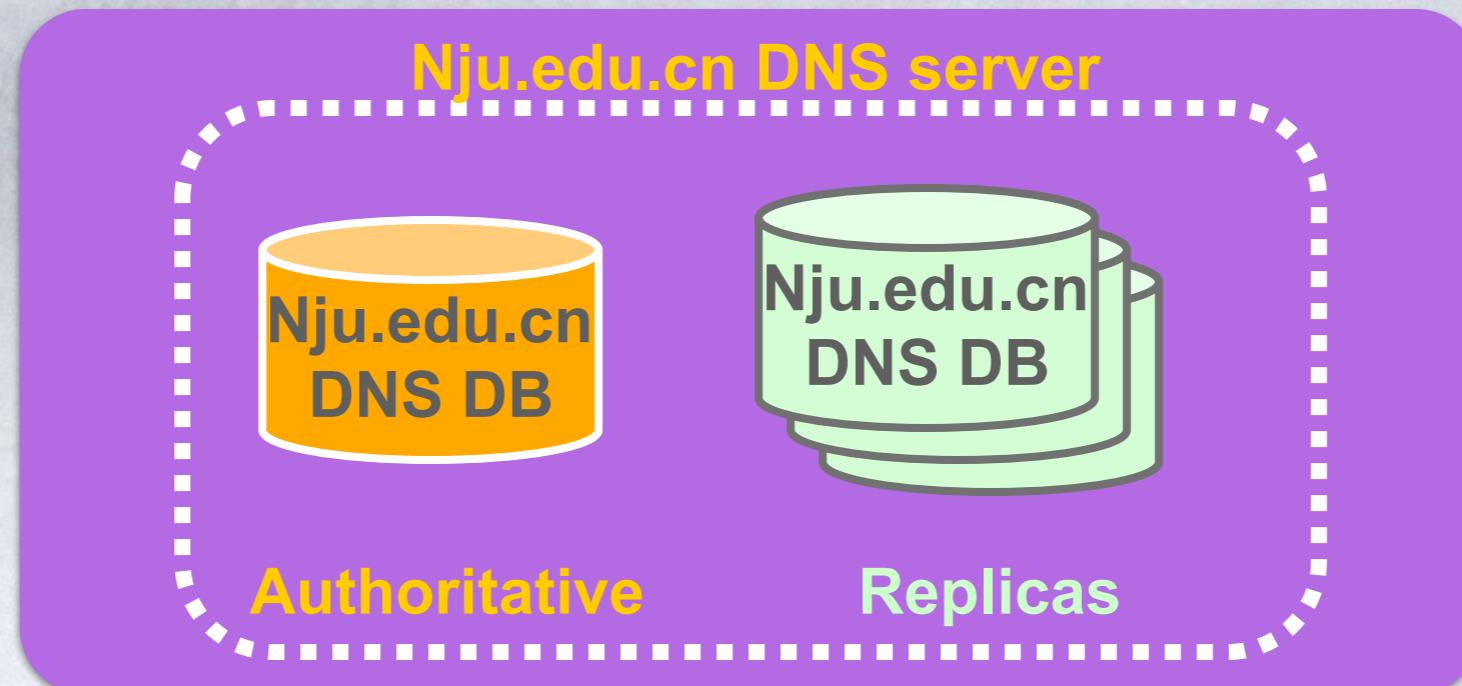
- ❖ Zone is a part of the subtree

- ❖ DNS server answers queries about hosts in its zone



Primary and secondary name servers

- ❖ For each zone, there must be a primary name server and a secondary name server
 - * The primary server (master server) maintains a zone file which has information about the zone. Updates are made to the primary server
 - * The secondary server copies data stored at the primary server.
- ❖ Adding a host:
 - * When a new host is added (“gold.cs.xxx.edu.cn”) to a zone, the administrator adds the IP information on the host (IP address and name) to a configuration file on the primary server

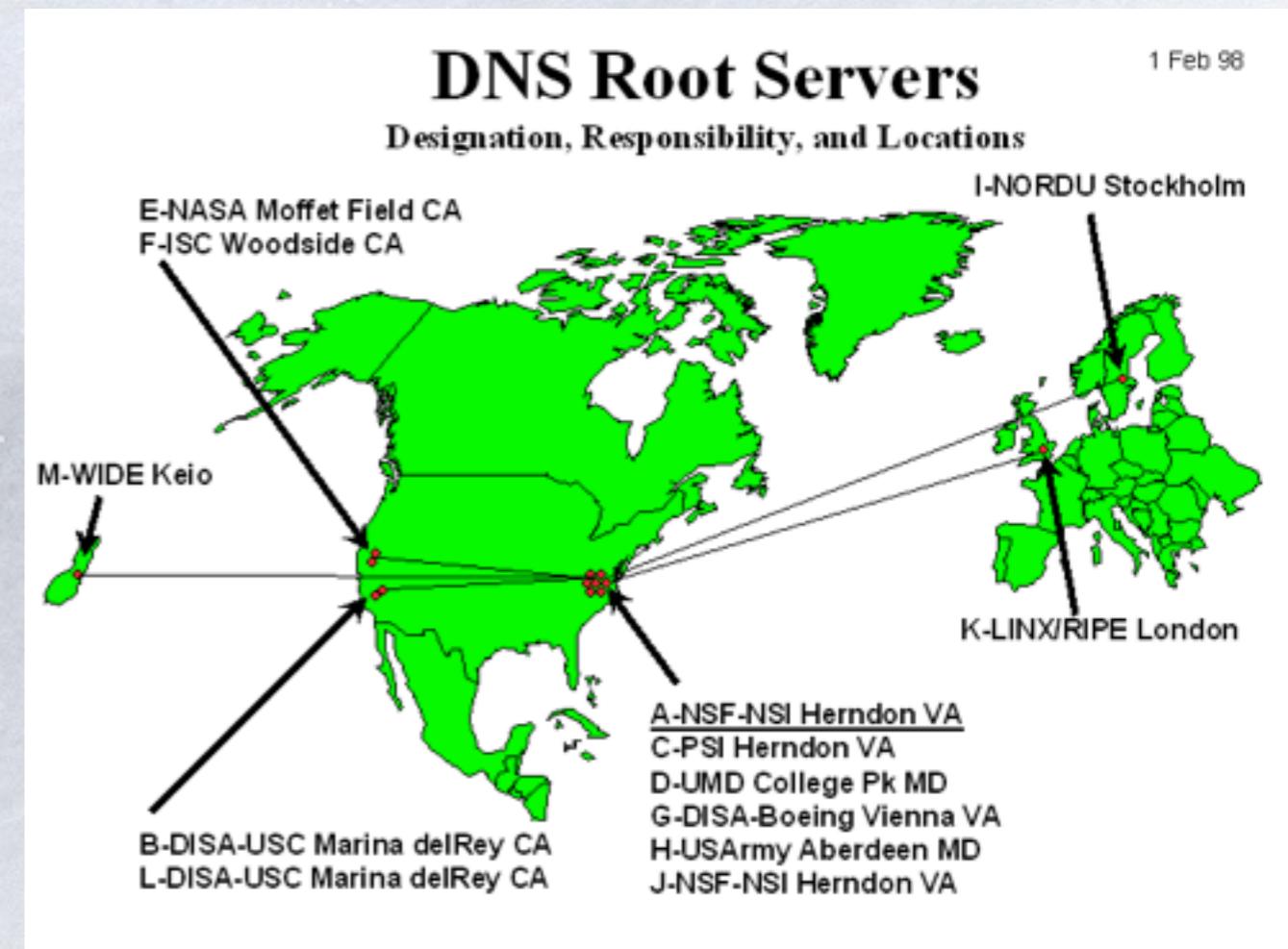


Root name servers

❄ The root name servers know how to find the authoritative name servers for all top-level zones.

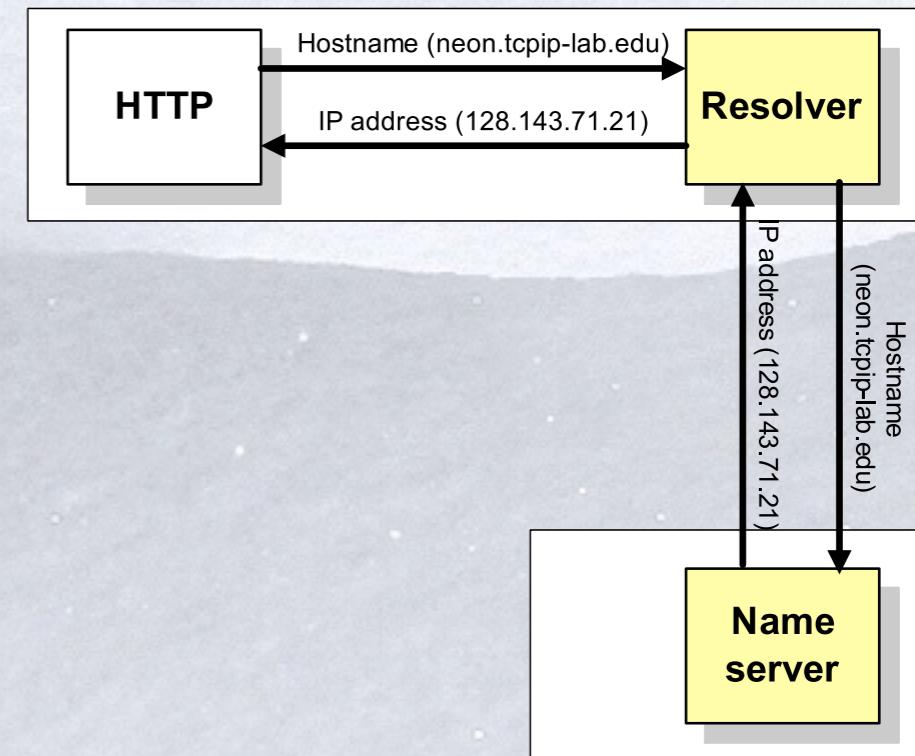
❄ There are only 13 root name servers

❄ Root servers are critical for the proper functioning of name resolution



Domain name resolution

- ❄ User program issues a request for the IP address of a hostname
- ❄ Local resolver formulates a DNS query to the name server of the host
- ❄ Name server checks if it is authorized to answer the query.
 - * If yes, it responds.
 - * Otherwise, it will query other name servers, starting at the root tree
- ❄ When the name server has the answer it sends it to the resolver.



Recursive and Iterative Queries

- ❖ There are two types of queries:
 - * Recursive queries
 - * Iterative (non-recursive) queries
- ❖ The type of query is determined by a bit in the DNS query
- ❖ Recursive query: When the name server of a host cannot resolve a query, the server issues a query to resolve the query
- ❖ Iterative queries: When the name server of a host cannot resolve a query, it sends a referral to another server to the resolver

Caching

- ❖ To reduce DNS traffic, name servers caches information on domain name/IP address mappings
- ❖ When an entry for a query is in the cache, the server does not contact other servers

Resource Records

- ❖ The database records of the distributed data base are called resource records (RR)
- ❖ Resource records are stored in configuration files (zone files) at name servers.
- ❖ DNS databases contain more than just hostname-to-address records:
 - * Name server records NS
 - * Hostname aliases CNAME
 - * Mail Exchangers MX
 - * Host Information HINFO

Resource Identifiers

❖ URI (Uniform Resource Identifier)

- * permits resources to reside anywhere on the Internet

❖ URL (Uniform Resource Locator)

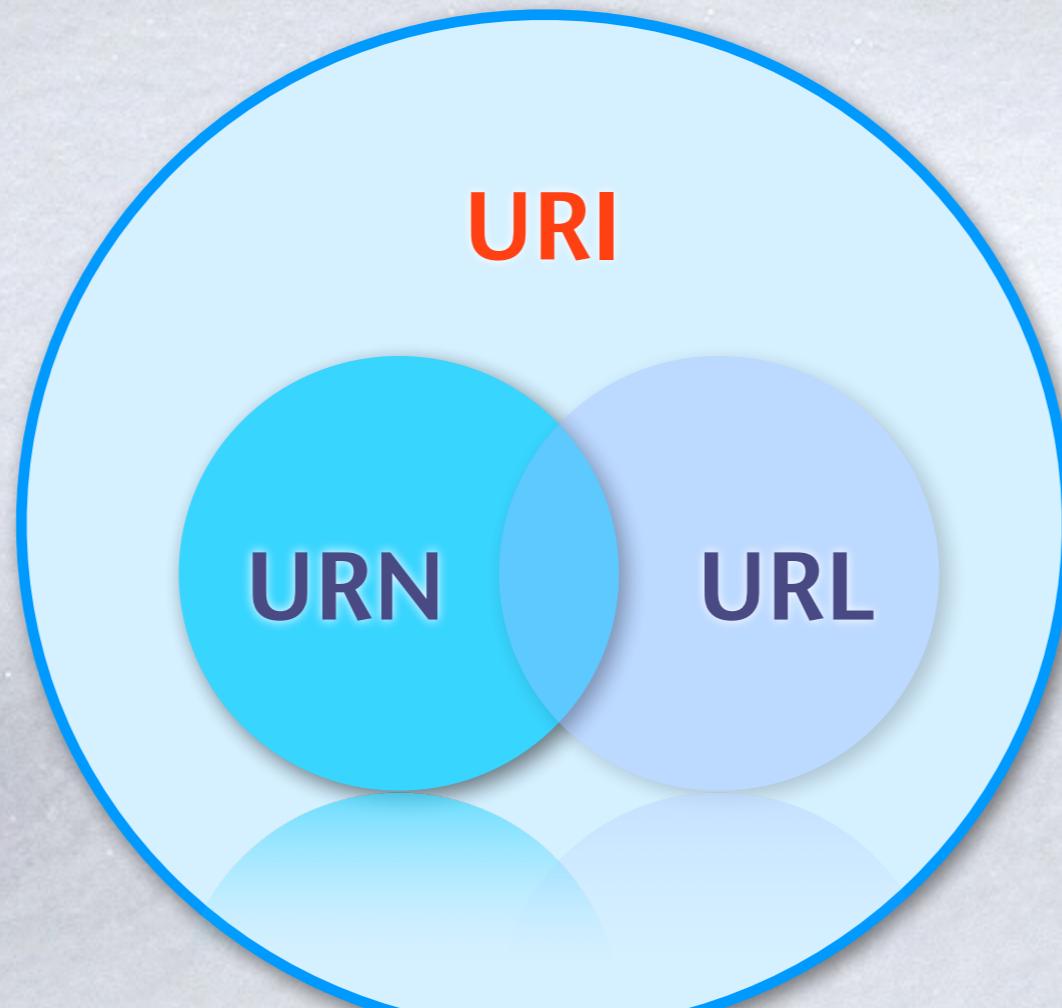
- * shows the location of a copy of a resource

❖ URN (Uniform Resource Name)

- * is a unique name for a resource

URI

- ❄️ URI is a superset of both URL and URN



URI: Universal Resource Identifier

❖ URIs defined in RFC 2396.

❖ Absolute URI:

- * scheme://hostname[:port]/path
- * http://www.nju.edu.cn:80/nic

❖ Relative URI: /path

- * /nic

- No server mentioned

URL

* Today, almost every URI is a URL

- * Most URLs follow a standardized format of three main parts:
 - * Scheme
 - * Server Internet address
 - * A resource on the web server

* more general format URL syntax:

scheme://domain:port/path?query_string#fragment_id

* Some example URLs

- * <http://www.joes-hardware.com/inventory-check.cgi?item=12731>
- * <ftp://joe:tools4u@ftp.joes.com/locking.gif>
- * <mailto:mac@yahoo.com>
- * rtsp://www.abc.com:554/interview/cto_video

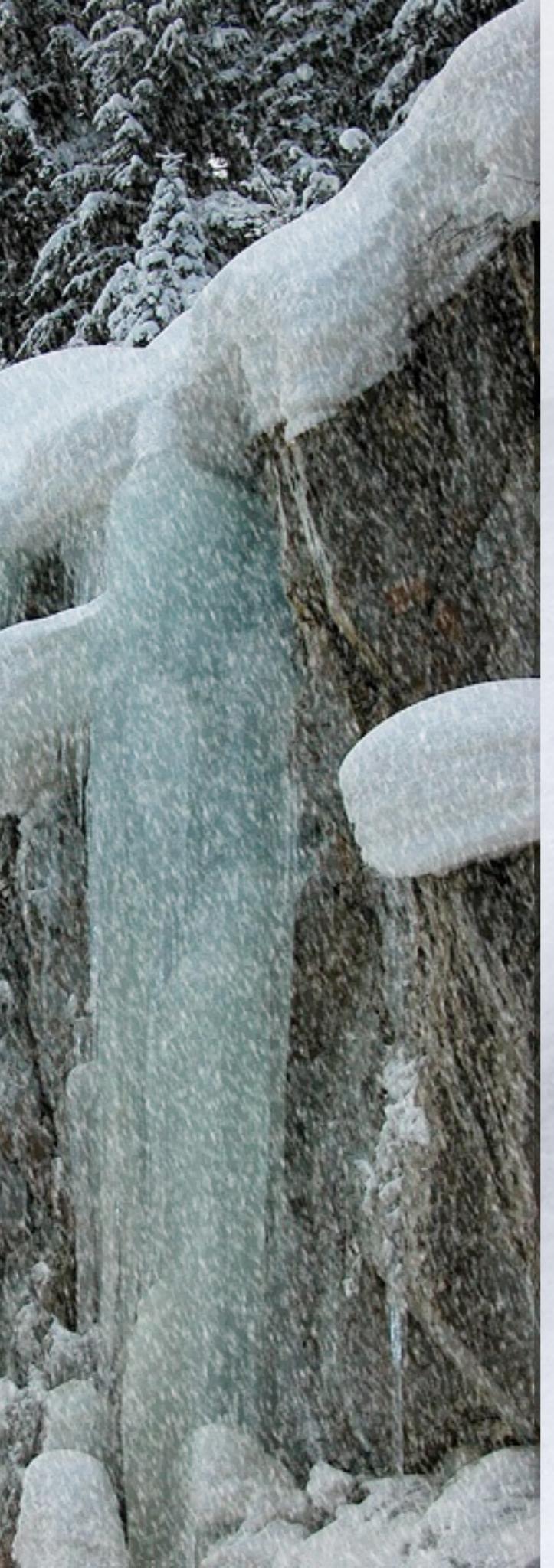
More advanced URLs

- ❖ **anchor:** jumps to a given section of a web page <http://www.textpad.com/download/index.html#downloads>
 - * fetches index.html then jumps down to part of the page labeled downloads
- ❖ **port:** for web servers on ports other than the default 80 <http://www.cs.xxx.edu:8080/secret/money.txt>
- ❖ **query string:** a set of parameters passed to a web program <http://www.google.com/search?q=miserable+failure&start=10>
 - * parameter q is set to "miserable+failure"
 - * parameter start is set to 10

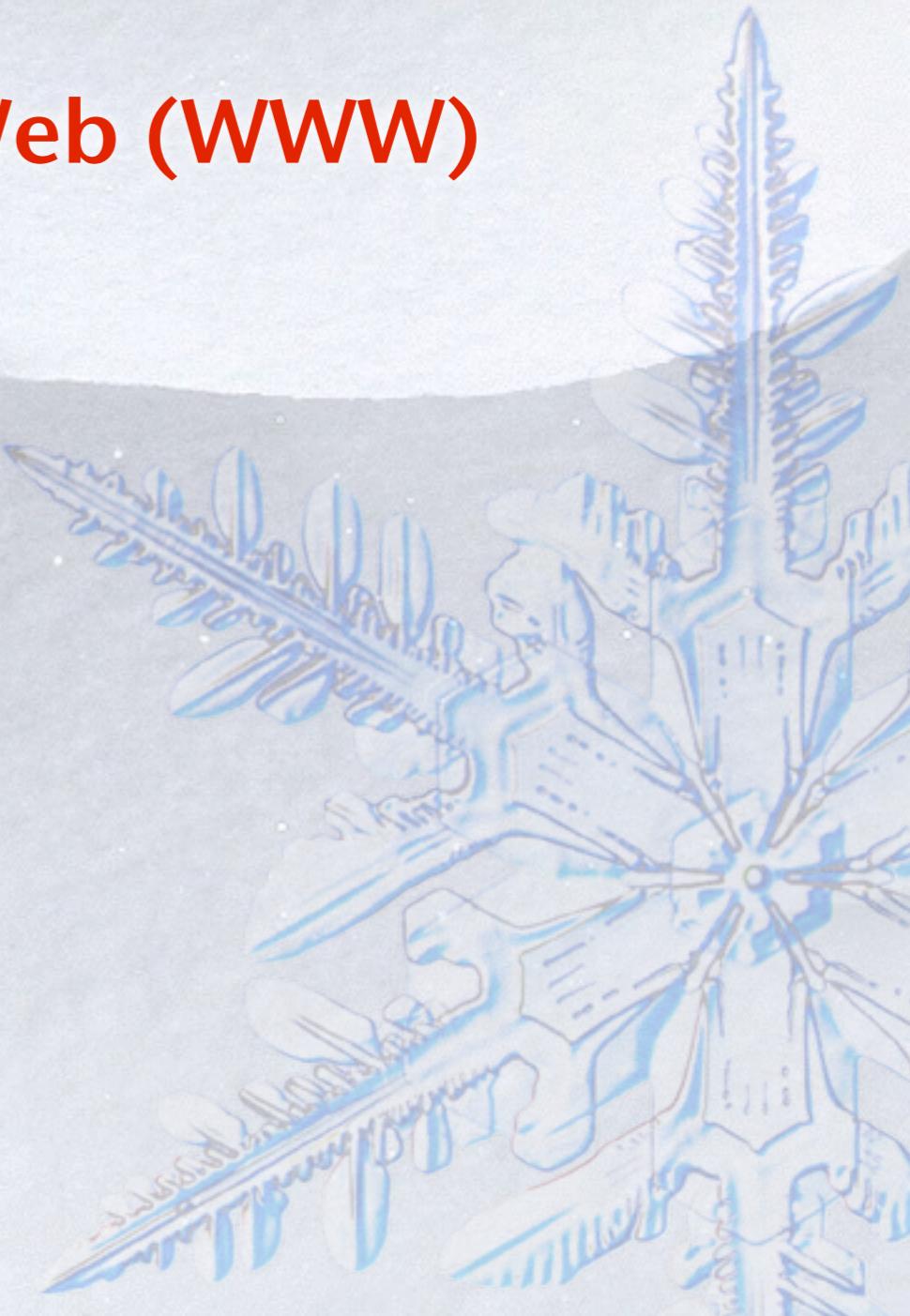
URN

- ❖ URN (uniform resource name)
 - * these location-independent URNs allow resources to move from place to place
- ❖ URN need a supporting infrastructure to resolve resource locations, that's why it is not yet widely adopted.
- ❖ Example URN

`urn:ietf:rfc:2141`



Outline

- ❖ The Internet
 - ❖ The World Wide Web (WWW)
- 

Web servers and browsers

❖ Web server: software that listens for Web page requests

- * Apache
- * Microsoft Internet Information Server (IIS) (part of Windows)

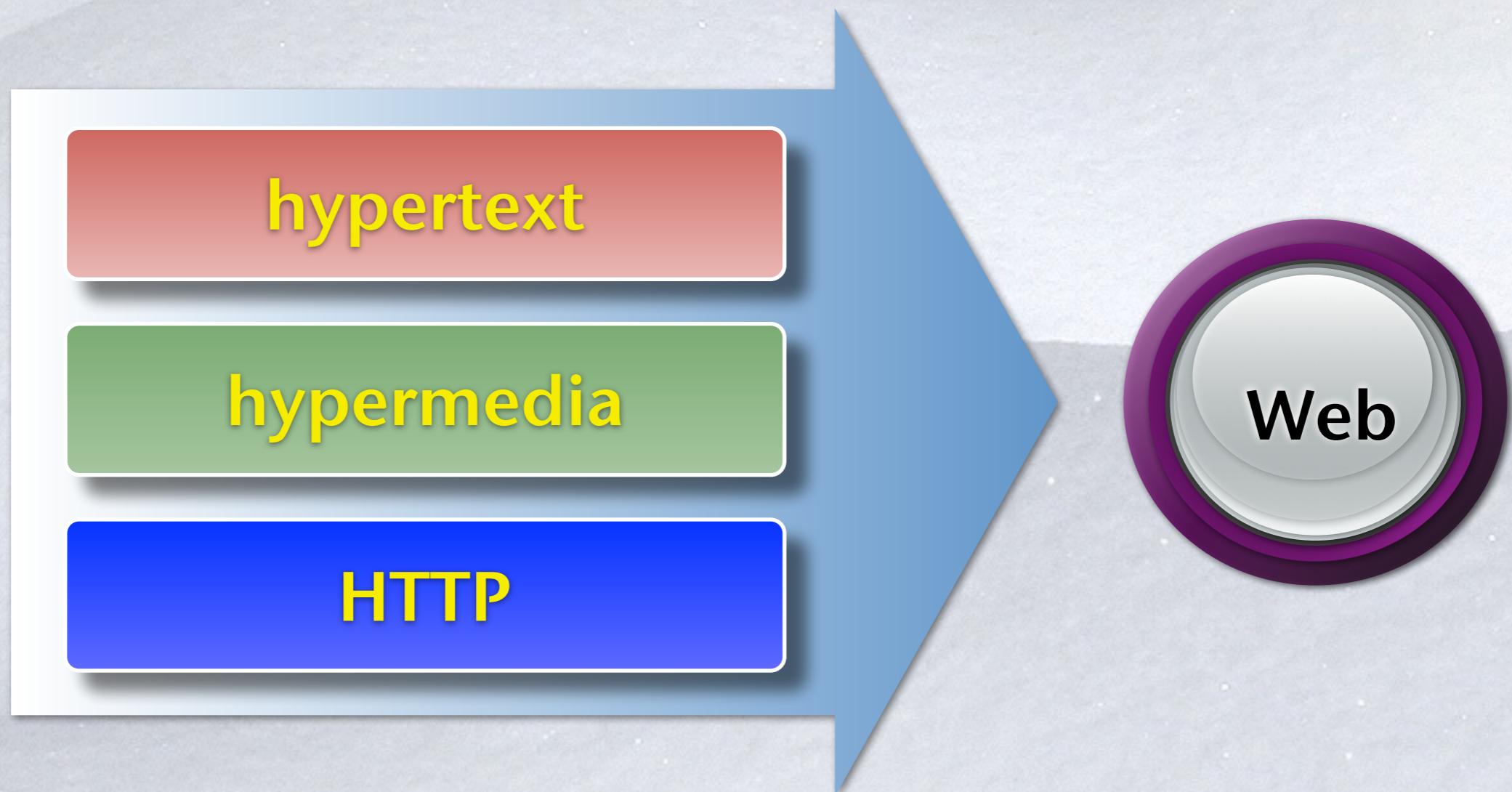


❖ Web browser: fetches/displays documents from Web servers

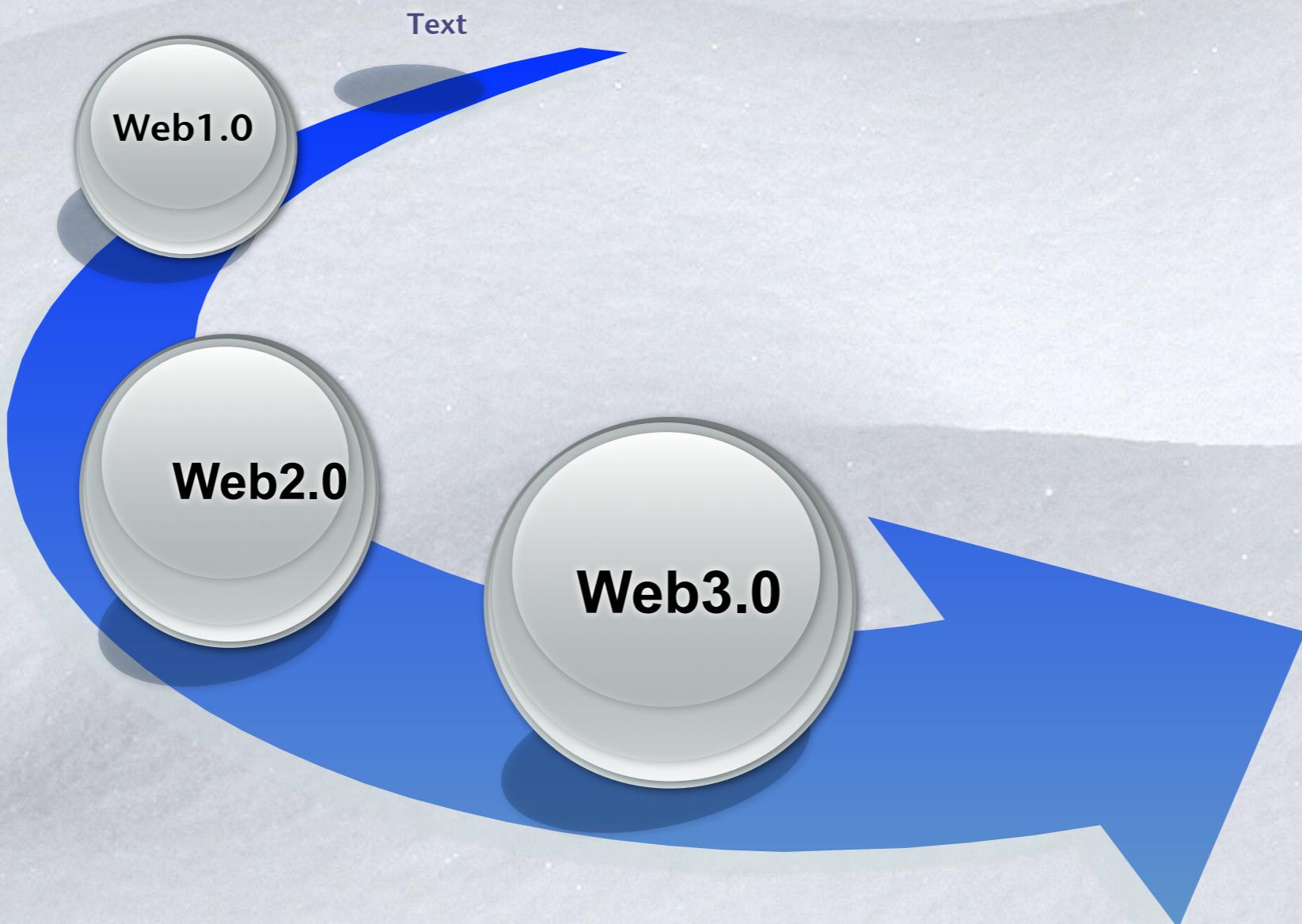
- * Microsoft Internet Explorer (IE)
- * Mozilla Firefox
- * Apple Safari
- * Google Chrome
- * Opera



Web的表现形式

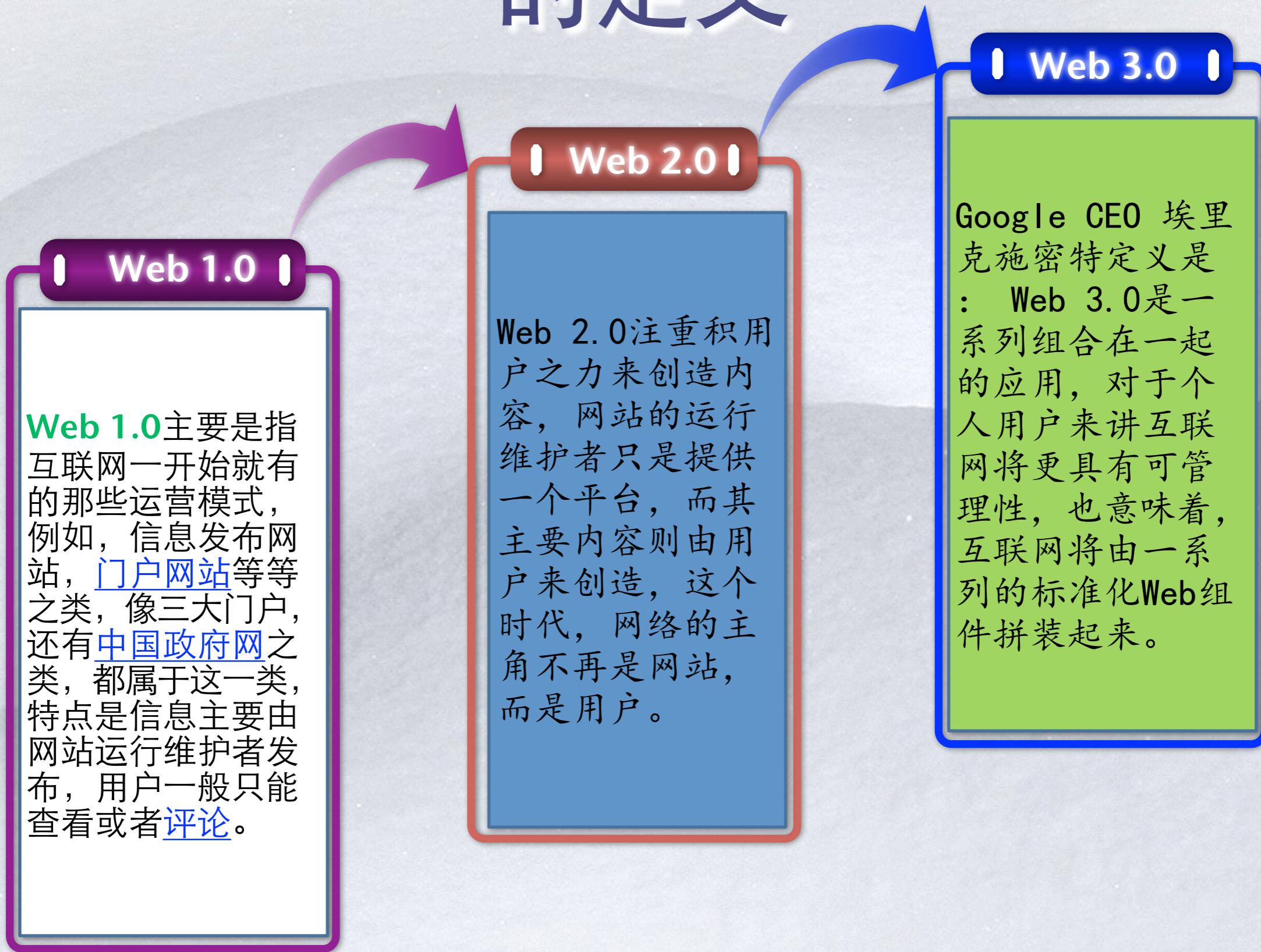


Web发展历程

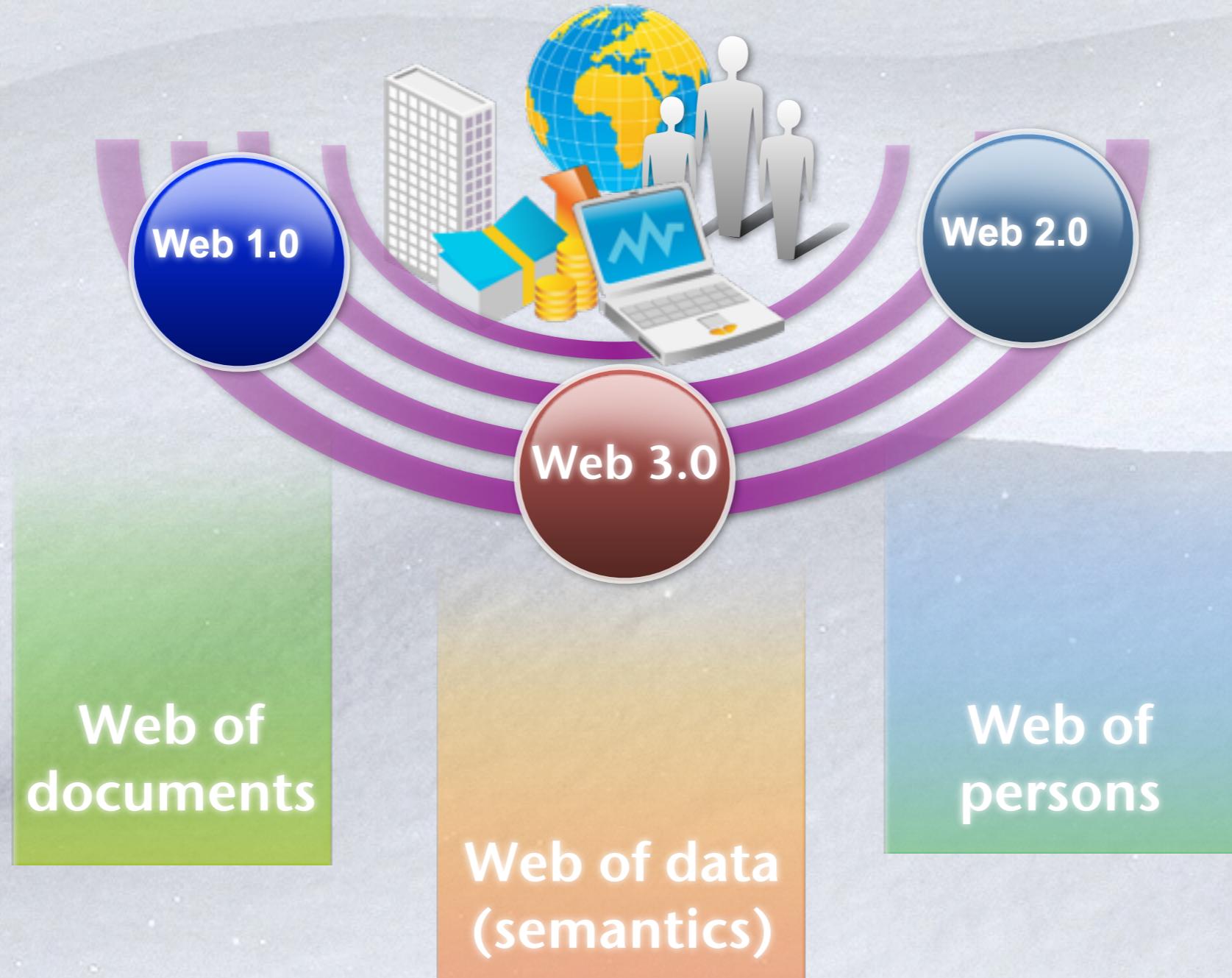


Web1.0 Web2.0 Web 3.0

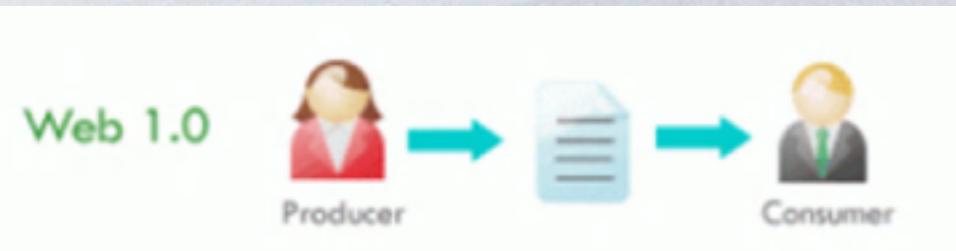
的定义



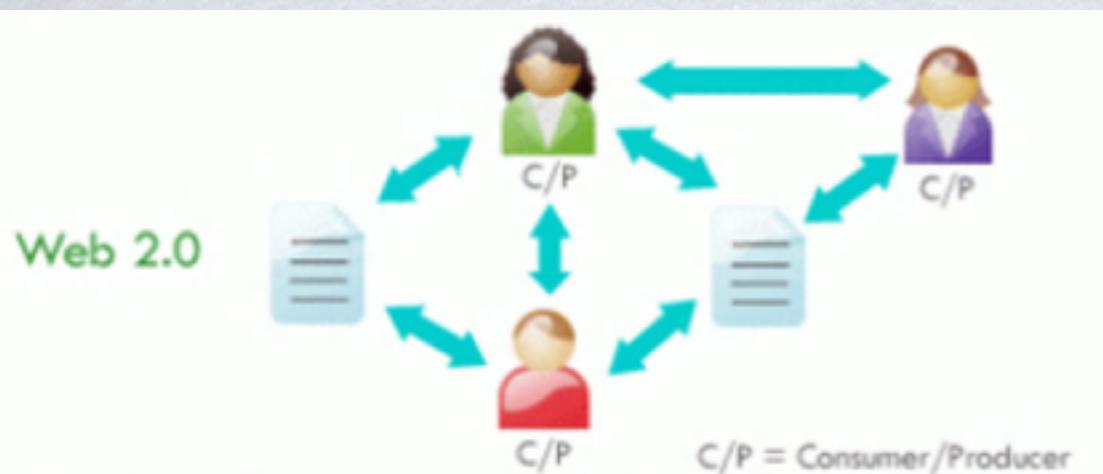
Web1.0 Web2.0 Web 3.0



Web1.0 Web2.0 Web 3.0 的比较



Web1.0: 网络资源从信息产生者到信息客户的单向流程



Web2.0: 所有人既是网络资源的信息产生者又是信息客户



Web3.0(语义网): 智能语义程序介入网络资源流程，更有针对性地发送信息和获得信息。

Web 2.0 examples

Web 1.0	Web 2.0
DoubleClick	→ Google AdSense
Ofoto	→ Flickr
Akamai	→ BitTorrent
mp3.com	→ Napster
Britannica Online	→ Wikipedia
personal websites	→ blogging
evite	→ upcoming.org and EVDB
domain name speculation	→ search engine optimization
page views	→ cost per click
screen scraping	→ web services
publishing	→ participation
content management systems	→ wikis
directories (taxonomy)	→ tagging ("folksonomy")
stickiness	→ syndication

Web 2.0 technologies

❄ browser side

- * Asynchronous JavaScript and XML (Ajax),
- * RIA
 - * Adobe Flash
 - * JavaScript/Ajax frameworks
 - * Prototype, script.aculo.us, Yahoo! UI Library, Dojo Toolkit, MooTools, jQuery, ExtJS, ...
 - * others
 - * XUL, JavaFX, Silverlight, OpenLaszlo, ...

❄ server side

- * many of same technologies as Web 1.0
 - * PHP, Ruby, ColdFusion, Perl, Python, JSP, Servlet, and ASP
- * addition with providing data in different format
 - * XML, RSS, and JSON , why?

Reading materials

- ❄ A Brief History of the Internet <http://www.isoc.org/internet/history/brief.shtml>
- ❄ http://en.wikipedia.org/wiki/Web_2.0
- ❄ <http://oreilly.com/web2/archive/what-is-web-20.html>

Thanks!!!

