



WWW-Administration

(*webmaster@lrz.de*)

September 2005

Folien :

`http://www.lrz.de/services/netzdienste/www/unixadmin_kurs/`

Was ist WWW?



Internet

Einige Hundert Millionen Rechner und Netzgeräte verbunden mit Leitungen oder Funk

Dienste im Internet

- Email, WWW
- DNS, NTP, Dateitransfer, Rechnerzugang
- Telephonie, Radio, Fernsehen, Videokonferenz, ...

WWW- World Wide Web

- Allgemeiner Informationsaustausch über Internet
- Integration von Text, Bild, Film und Ton
- **Links, Interaktivität**
- Client-Server-Konzept

- 1989** Tim Berners-Lee: Vorschlag für CERN-Management
- 1991** Erste WWW-Version von CERN veröffentlicht:
HTTP, HTML, URL
- 1993** Mosaic-Browser von Marc Andreessen et.al. (NCSA)
ca. 50 bekannte HTTP-Server
- 1995** Start des Apache-Projekt; ca. 23.500 sites

.....

HTTP 1.0 → HTTP 1.1 (1999)

HTML → HTML 4.01 (1997), CSS, XHTML, XML

Websites → Webportale

Suchmaschinen

E-Commerce, E-Government, E-Banking

Webserver allüberall

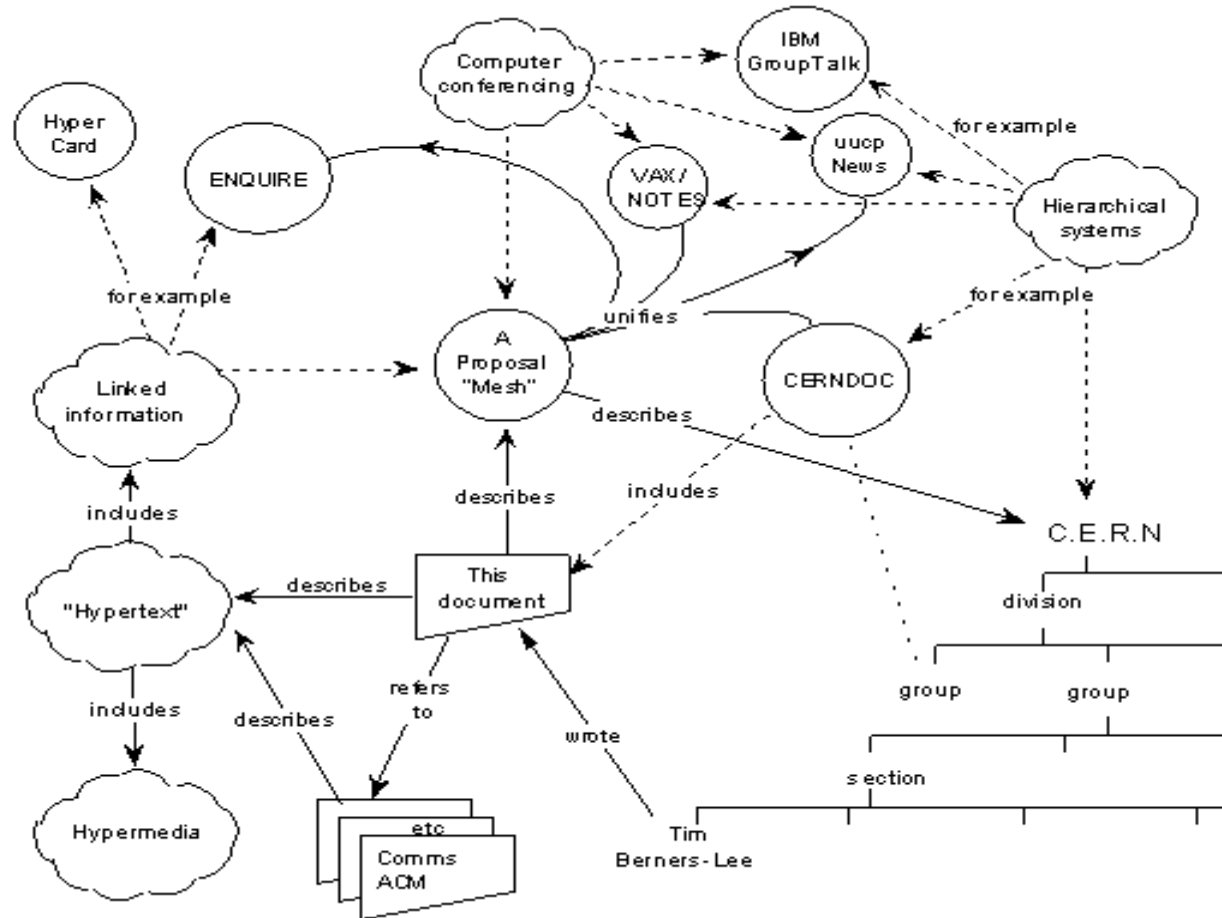
.....

2005 Ca. 58 Mio. sites (Januar 2005)

Geburtsurkunde des WWW 1989



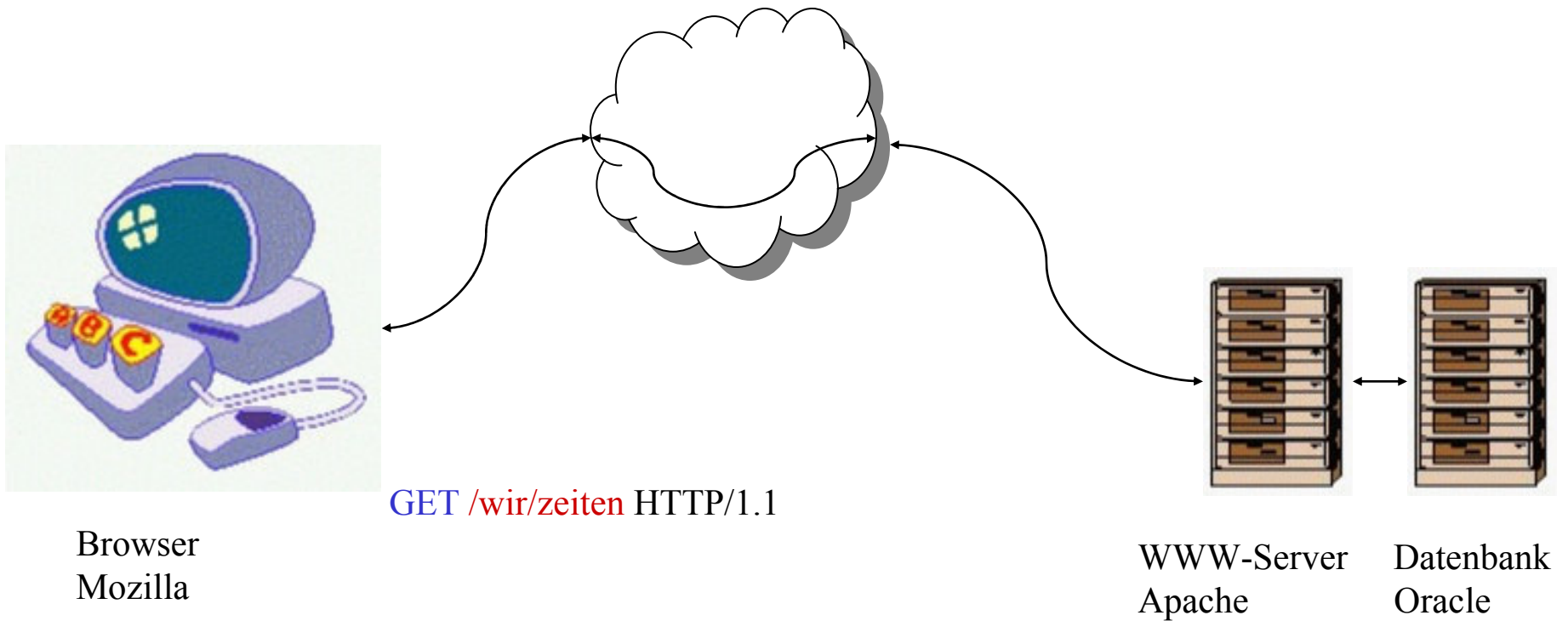
Information Management: A Proposal (Tim Berners-Lee / CERN)



Was macht ein WWW-Server?



URL: <http://www.lrz-muenchen.de:80/wir/zeiten>



DNS-Namen festlegen

Dienstnamen statt Rechnernamen

Kurz-/Langnamen: `www.lmu.de` <-> `www.uni-muenchen.de`

Dienste

Statische Inhalte (HTML), Dynamische Inhalte (PHP, CGI, Java, ASP)

Portal, Suchmaschine, Webmail, ...

Nutzung

Verbreitung: weltweit / firmenintern

Zugang: öffentlich / beschränkt

Betriebsmittel

Wieviel Maschinenleistung, wieviel Speicherplatz, welche Netzanbindung, Ausfallsicherheit (Dauerbetrieb!)

Datenquellen

Eigene Autoren, Dritte, automatische Daten

	Inhalt	Gestaltung	Technik
Alles selbst	x	x	x
ISP (z.B. LRZ)	x	x	-
ISP + Webdesigner	x	-	-

Inhalt: Fachleute, Juristen (z.B. Impressumspflicht, Rechte Dritter)

Gestaltung: Graphik, Standards, Navigation, Barrierefreiheit, Rechtsfragen, HTML, PHP-, CGI-, ... Programmierung

Technik: Netzanbindung, Webmaschinen, Webprogramme, Datenbanken, Datensicherung

WWW-Server	Apache, SUN ONE, Internet Information Server (IIS)
Module zum WWW-Server	Verschlüsselung: SSL Anbindung: Skriptsprachen, Application Server Validierung: LDAP, AFS, Novell eDirectory
Skriptsprachen	PHP, Perl, Python
Application Server	Tomcat, ServletExec, Zope
Datenbanken	Oracle, MySQL, PostgreSQL
Suchmaschine	mnoGoSearch, ht://Dig, Harvest, Fast, Google-Appliance
Gestaltung	Dreamweaver, GoLive, Frontpage
Content Management	Zope+Plone, NPS (?), Communiqué, Typo3, LRZ-Publish

Zentrale Konfigurationsdatei: `httpd.conf`

Gelesen: Start des Dämons, bzw. Signal an Dämon

Enthält: Alle möglichen Apache-Anweisungen

Setzungen für Teilbereiche:

```
<Directory ...> ... </Directory>
```

```
<Location ...> ... </Location>
```

```
<VirtualHost ...> ... </VirtualHost>
```

Erlaubnis für dezentrale Konfiguration:

```
AllowOverride All|None|Authconfig|...|Options
```

Dezentrale Konfigurationsdateien: `.htaccess`

Gelesen: Bei jeder Anfrage in jedem Ordner oberhalb Zieldatei

Enthält: Viele Apache-Anweisungen möglich

ServerType	standalone
KeepAlive	TCP/IP-Verbindung halten (On/Off)
MaxClients	Maximale Zahl Serverprozesse
Listen	Portnummern: HTTP=80, HTTPS=443
LoadModule	Module bei DSO-Variante (optional)
User	Kennung Serverprozess (Nie unter root!)
Group	Gruppe Serverprozess (Nie unter root!)
ServerAdmin	Kontaktadresse (E-Mail)
ServerName	DNS-Name, z.B. <code>www.lrz-muenchen.de</code>
DocumentRoot	Pfad zu Nutzdaten
DirectoryIndex	<code>index.html index.htm index.php</code>
ErrorLog	Pfad Fehler-Log
CustomLog	Pfad Zugriffs-Log

Aliasnamen

```
Alias /image /ftp/pub/image
```

CGI-Programme

```
ScriptAlias /cgi-bin/ /local/webserver/cgi-bin/
```

URLs umschreiben

```
RewriteEngine on
```

```
RewriteRule ^/(.*) http://www.lrz-muenchen.de/$1 [R]
```

Zugriffsschutz (**Persönliche Passwörter über HTTPS!**)

```
AuthType Basic; AuthName Private; AuthUserFile ...;  
Require valid-user
```

Aktivierung PHP

```
AddType application/x-httpd-php .php
```

Viel mehr: <http://httpd.apache.org/docs/mod/directives.html>

Content Management System: Warum?



Übersicht

- Wer, Wann, Wieviel?

Leichtere Eingabe

- Vorlagen, WYSIWYG-Editor, Konverter, reale Vorschau

Einheitliche Gestaltung

- Seitenaufbau, Logos, Impressum, Farben, Navigation, ...

Rechteverwaltung

- Wer darf was? Zugriffskonflikte lösen

Versionierung

- Alte + zukünftige Varianten, Life + Produktionssystem

Zeitgesteuerte Veröffentlichung

- Automatische Veröffentlichung / Löschung

Trennung Form und Inhalt

- Leichte Änderbarkeit
- Mehrere Ausgabemedien: WWW, PDF, E-Mail, ...

Auto-Content

- Navigationselemente, Sitemaps, Inhaltsverzeichnisse

Workflow

- Freigabe durch Vorgesetzte, mehrere Autoren

Link-Management

- Problem: tote Links nach Löschung von Dokumenten

Mehrsprachigkeit

- Webseiten in mehreren Sprachen
- Mehrsprachige Benutzerführung

Mandantenfähigkeit

- Mehrere Webserver mit einem CMS

Suchmaschine

- Suchen mit Zusatzinfos aus CMS

Personalisierung

- Zugriffsschutz, personalisierte Inhalte



1. HTML-Editoren mit Upload/Download-Funktion

- Nur Dokumente, keine Datenbanken

2. Redaktionssysteme

- Einige Teile von Webseiten dynamisch aus Datenbanken

3. (W)CMS – Web Content Management System

- Komplette Website verwaltet, viel aus Datenbanken

4. ECMS – Enterprise Content Management Systems

- Integration vieler Daten und Anwendungen aus der Organisation

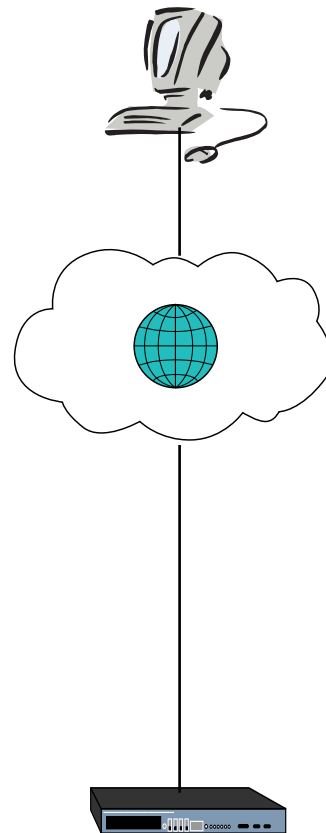
Webfarm: Schema



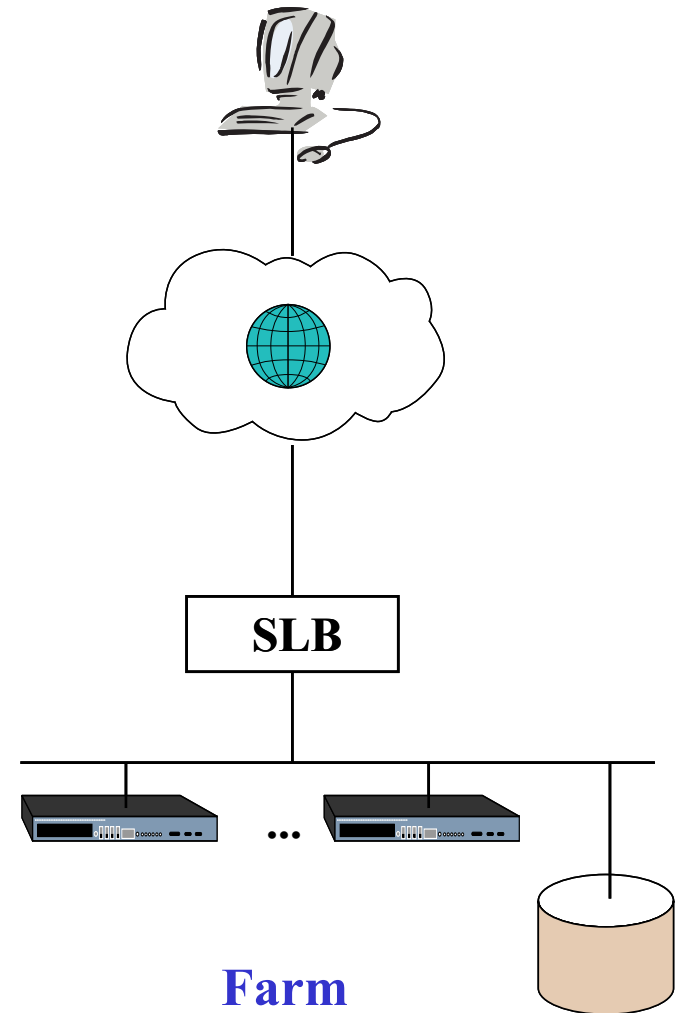
WWW-Client

Internet

Webmaschinen

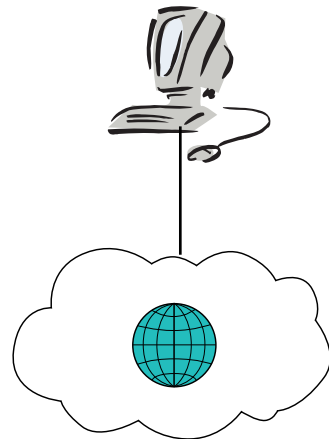


Einfach



Farm

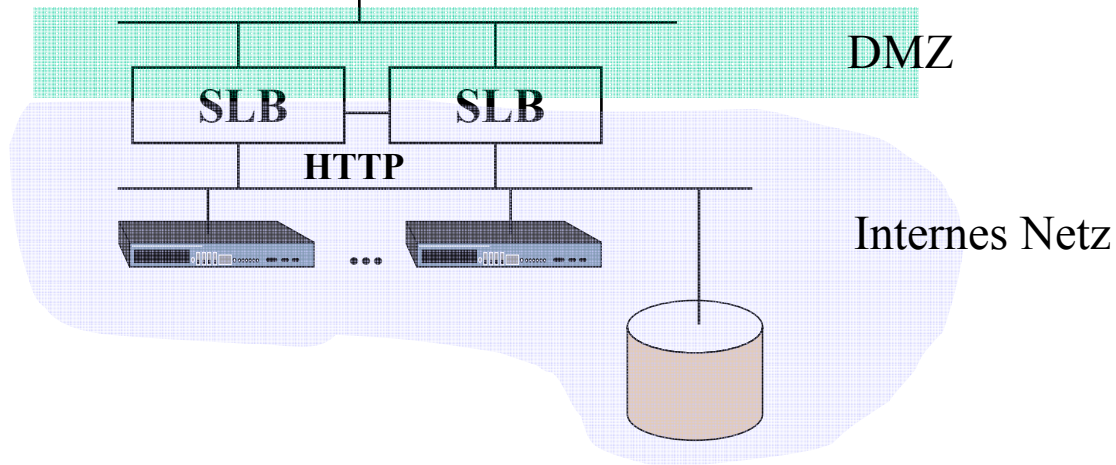
WWW-Client



Internet

Server Load Balancer (SLB):
Network Address Translation (NAT)
Lastverteilung
SSL-Beschleunigung
URL-Umschreiben

HTTP HTTPS



Webmaschinen

Gefahr durch CGI

- Laufen unter Kennung des Servers
Jeder CGI-Autor kann alle Server-Daten überschreiben
- Nie fremde CGIs bedenkenlos verwenden
- Nur einzelne Bereiche für CGI erlauben

Gefahr durch UNIX-Links

- Link umgeht geschützte Bereiche (`.htaccess`)
- Zugriffsrechte auch für Unterverzeichnisse setzen

Verschlüsselung Passwort

- Bei HTTP-basic-authentication nur `uuencode`

Verschlüsselung aller Daten mit geeigneten Protokollen

- SSL, TLS

Eigene Suchmaschine

- Programm: Harvest, ht://Dig u.v.a. Oder einfach **Google**.
- Festlegen des Suchbereichs
Dokument-Typen (z.B. keine Bilder), Dateibäume (Download)
- Gestaltung des Suchformulars
und zugehöriger CGI-Skripten

Zugriffsteuerung für fremde Suchmaschinen

- <http://www.my.domain/robots.txt>

```
User-agent: *  
Disallow: /cgi-bin/  
Disallow: /Harvest  
Disallow: /aktuell/www
```

- Meta-Information im HTML-Header

```
<META NAME=„robots“ CONTENT=„(no)index, (no)follow“>
```