

# Sdílení informací a dat

Ing. Michal Radecký, Ph.D.

# Sdílení souborů a zdrojů

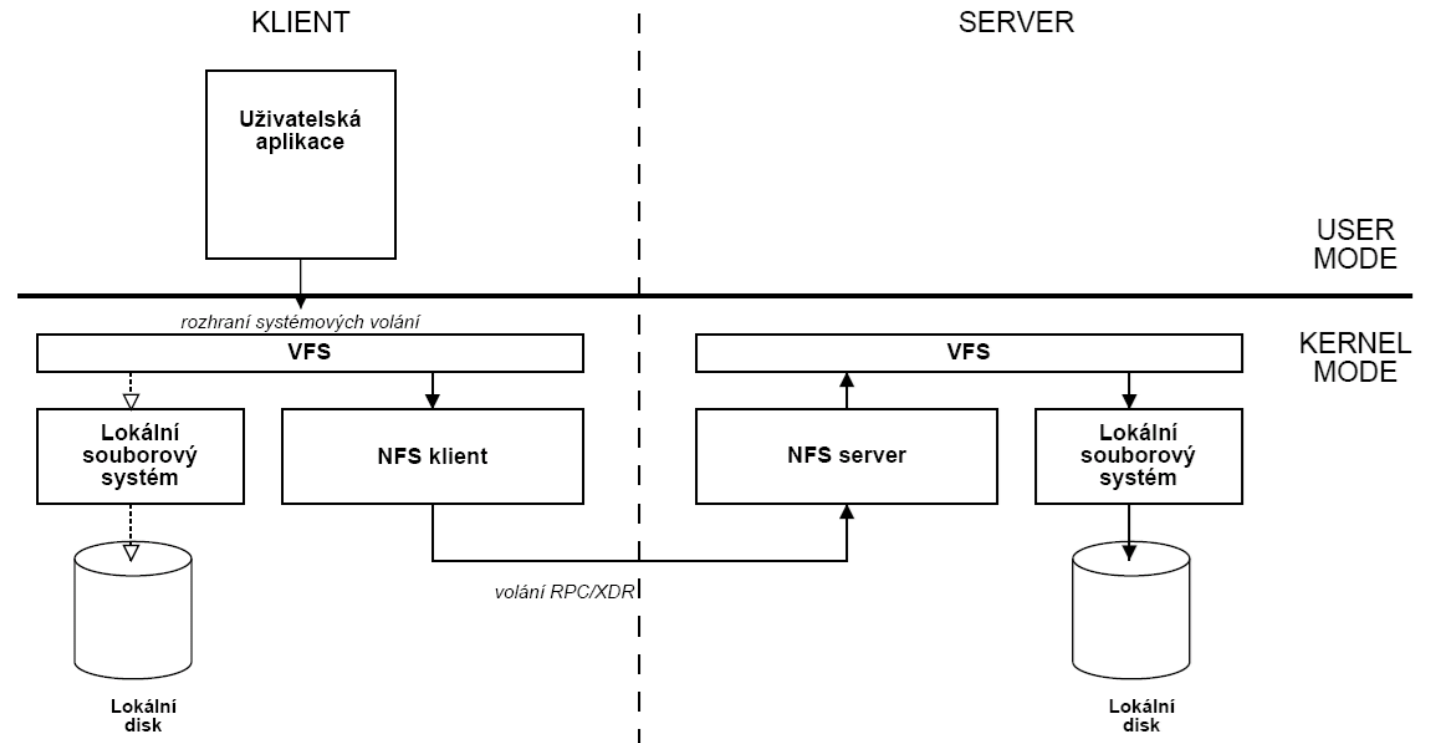
- Zpřístupnění zdrojů (soubory, tiskárny, apod.) různým klientům, a to v prostředí internetu nebo místních sítí.
- Různé pohledy na zajištění sdílení
  - HTTP (WWW, Webové služby, Stream, atd.)
  - FTP, SFTP, atd.
  - sdílené souborové systémy
  - výměnné systémy

# Technické přístupy přenosu

- Transparentní
  - sdílený soubor se „tváří“ jako lokální. Veškeré operace nutné pro zajištění přístupu jsou řešeny skrytě před uživatelem (NFS protokol)
- Netransparentní
  - sdílený soubor je nutné chápat jako vzdálený, je nutné jej stáhnout, aby mohl být využíván jako lokální (FTP protokol)
- Stavový
  - oba konce komunikace udržují svůj vnitřní stav v závislosti na procesu komunikace, kdy musí být tyto stavy ve vzájemném souladu podle procesu komunikace (TCP protokol)
- Bezstavový
  - neexistuje žádná stavová informace, server je po přijetí a zpracování požadavku opět ve stejném stavu. Nižší nároky na spolehlivost přenosu (UDP protokol)
- Symetrický
  - počítače vystupují jak v roli serveru, tak klienta – základní prvek obecných P2P sítí (SMB protokol)
- Asymetrický
  - vždy existuje server nabízející služby a klienti, kteří tyto služby využívají. Toto rozdělení rolí se nemění (HTTP protokol)

# NFS (Network File System)

- Veřejně představeno v roce 1989 firmou Sun Microsystems
- Aktuálně verze 4 (RFC 3530 pod IETF)
- Transparentní protokol, původně bezstavový, dnes již poskytuje také stavový přístup
- Využívá RPC (Remote Procedure Call)
- Pracuje pevně na portu 2049
- Podpora bezpečnosti, zamykání souborů, cachování, atd.
- WebNFS – rozšíření pro lepší integraci přímo do webových prohlížečů a pro průchodnost sítěmi



# SMB/CIFS (Server Message Blocks)

- Vývoj zahájen v roce 1984 firmou IBM, od roku 1986 pod hlavičkou Microsoft a Intel, dnes pouze Microsoft
- Označován jako standard CIFS (RFC 1001, 1002, atd.)
- Windows 10 využívají komunikační specifikaci SMB 4.0 (řadič Active Directory)
- Dnes nejrozšířenější P2P protokol pro sdílení prostředků v lokální síti
- Opět využívá RPC
- Pracuje na portech 445 (TCP/IP), 137-139 (NetBIOS)
- NetBIOS je rozhraní/protokol využívající logická jména pro adresování, nezávislost na nižších vrstvách

ISO/OSI			TCP/IP		
Aplikační	SMB			Aplikační	
Presentační					
Relační	NetBIOS	NetBEUI	NetBIOS	NetBIOS	TCP/UDP
Transportní	IPX		DECnet	TCP+UDP	
Síťová				IP	IP
Linková	802.2, 802.3, 802.5	802.2, 802.3, 802.5	Ethernet V2	Ethernet V2	Linková
Fyzická					

# Výměnné systémy

- Komunikační systémy postavené na principu P2P sítí, které ke svému chodu vyžadují speciální softwarové vybavení
- Primárně jsou určeny k výměně dat/souborů mezi jednotlivými uživateli, kdy poskytují služby na dvou vrstvách
  - vyhledání zdrojů
  - přenos zdrojů
- Základní vlastnosti
  - netransparentní
  - stavový i nestavový (v závislosti na transportním protokolu)
  - symetrický i nesymetrický (v závislosti na konkrétní technologii a fázích komunikace)
- Základní typy
  - Centralizované vyhledávání (DirectConnect)
  - Decentralizované vyhledávání (Gnutella)
  - Distribuované vyhledávání (FastTrack)
  - Nedistribuované stahování (Napster)
  - Semi-distribuované stahování (FastTrack)
  - Plně distribuované stahování (BitTorrent)

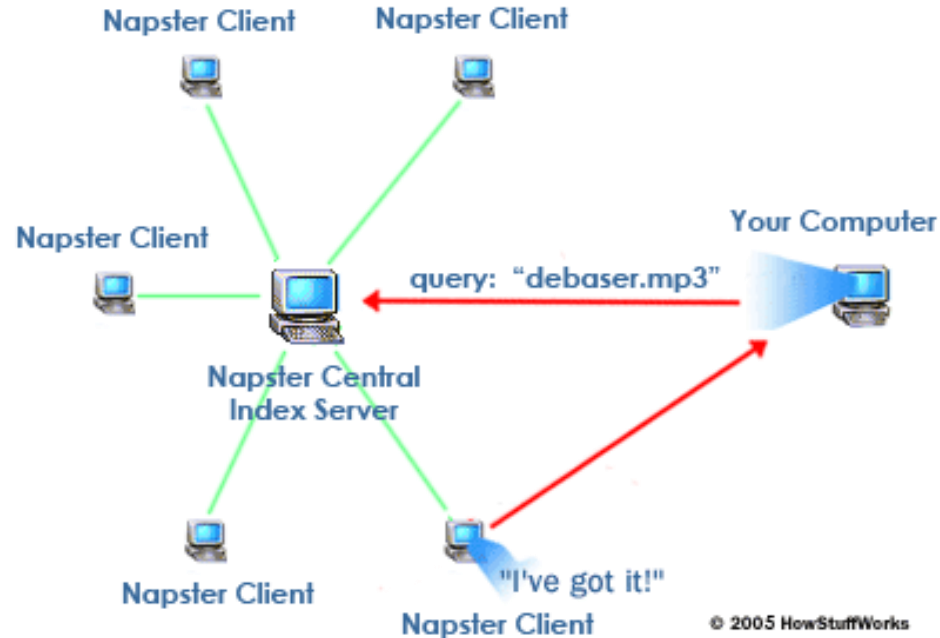
# Napster

- Shawn Fanning v roce 1999 začal vyvíjet protokol pro vyhledávání a sdílení MP3 souborů mezi svými přáteli
- První skutečný P2P systém pro sdílení souborů
- Nabízel také služby chatu (IRC)
- Na vrcholu (2001) registroval 30 mil. uživatelů a indexoval 2,8 miliardy souborů
- Pouze jedna klientská aplikace
- Vyhledávání
  - centrální server s aktuální databází uživatelů a jejich souborů – index, metasever
  - využíval protokol TCP (8875)
- Přenos souborů
  - navázání spojení s jedním vybraným poskytovatelem obsahu
  - stáhnutí souboru (jako celku) a ukončení komunikace
  - využíval protokol TCP (7777, 8888)

# Napster

- Problémy
  - rostoucí velikost indexového souboru
  - legálnost sdílení dat (friends-to-friends)
- Napster 2 – nyní provozovaná komerční verze pro sdílení hudby

## Napster Protocol



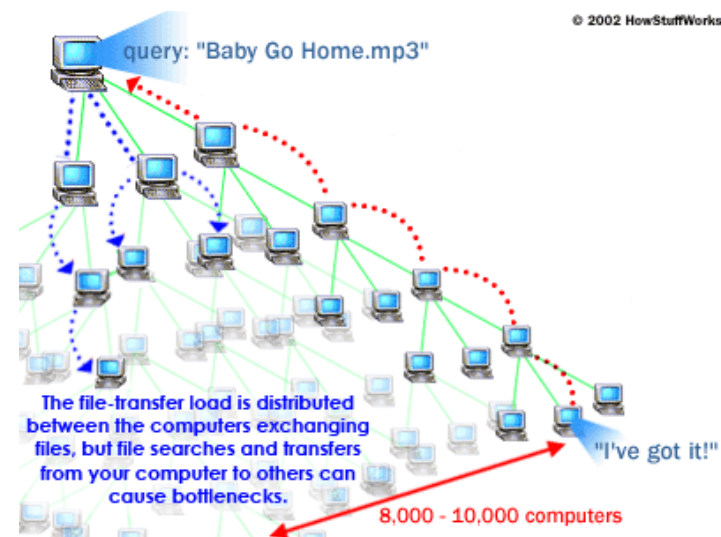


# Gnutella

- Nullsoft (autor WinAmp) v roce 2000 na svých stránkách zveřejnil aplikaci pro P2P, jako reakci na problémy Napsteru. Aplikace byla brzy stažena, nicméně princip sítě již existoval
- Jedná se o síť, na které funguje celá řada klientských aplikací, které přinášejí rozšířenou funkcionalitu nad hlavními službami sítě (stahování z více zdrojů, apod.)
- Řeší především nedostatky Napsteru
  - neexistuje centrální index, vše se řeší distribuovaně
  - jiný přístup k legalitě provozu díky distribuci vyhledávání a umístění dat
- Vyhledávání
  - dotaz je zaslán dalším počítačům, o kterých daný klient ví. Tito sousedé kontrolují svá lokální úložiště a v případě, že soubor mají, tazatele informují. V případě, že soubor nemají, opět oslovují své sousedy.
  - omezení pomocí TTL (Time-To-Live) dotazu
- Stahování
  - vytvoření spojení mezi tazatelem a počítačem, který má daný soubor k dispozici
  - stáhnutí souboru (jako celku nebo částečně) a ukončení komunikace
  - realizováno přímo, a to pomocí HTTP protokolu

# Gnutella

- Problémy
  - generování velkého provozu v síti a pomalé dotazování – distribuované dotazy
  - vzhledem k omezení hloubky dotazu (TTL) nemusí být hledaná data nalezena, i když v globálním pohledu v síti existují
  - využívání každého počítače jako uzlu pro zpracování jiných dotazů
  - leeching – poskytnutí počítače pouze na nezbytně dlouhou dobu s omezenou nabídkou ke sdílení
- Gnutella 2 – úprava komunikačních mechanismů a vyhledávacích algoritmů

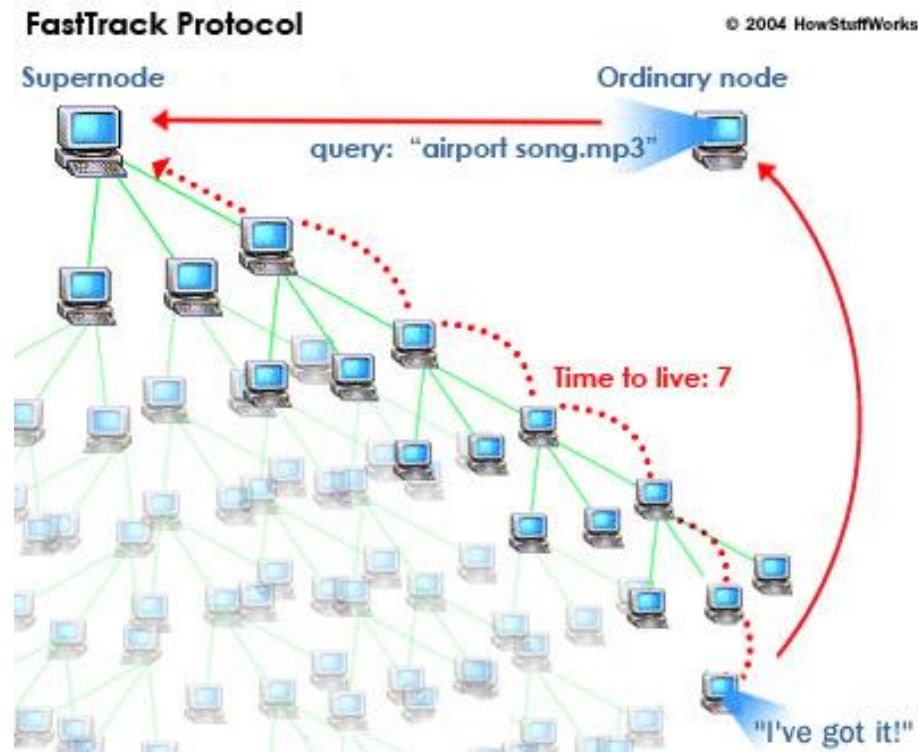


# FastTrack

- Zveřejněno v roce 2001 autory pozdějšího Skypu
- Synonymum Kazza
- Založeno na principu „supernodes“, kde je funkčnost podobná jako v případě Gnutelly, nicméně koncový uzel komunikuje pouze se svým primárním supernodem.
- Vyhledávání
  - klient se pokusí připojit k supernodům, které zná. Po připojení dostane seznam aktuálních supernodů a vybere si jeden primární. Tomu zašle informace o svých datech a dotaz k hledání.
  - supernod se postará (ve spolupráci s dalšími supernody) o nalezení úložiště hledaného souboru a zašle informaci zpět klientovi.
  - využívá TCP i UDP
- Stahování
  - na základě získaných informací dochází k přímému spojení pomocí HTTP protokolu
  - možnost částečného stahování souboru z různých zdrojů

# FastTrack

- Problémy
  - „hybridní“ architektura, kdy je potřeba zajistit provoz supernodů



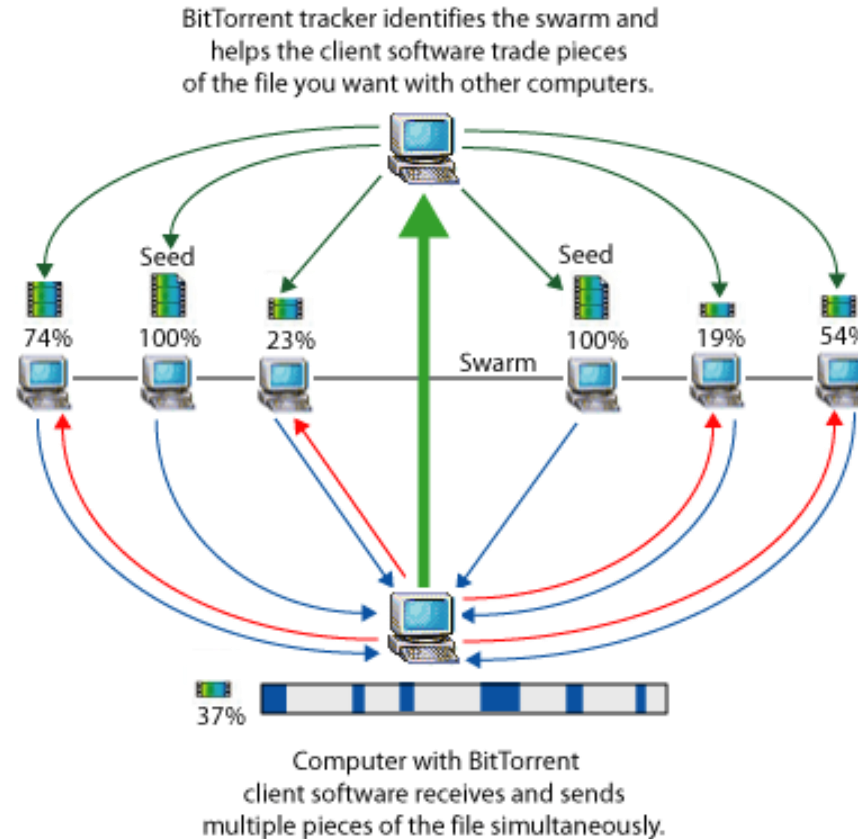
# BitTorrent

- Bram Cohen v roce 2002 uvedl aplikaci v Pythonu, která se stala původním klientem sítě BitTorrent pod Open Source licenci
- Původně určeno pro urychlení stahování velkých souborů, o které je jednorázový zájem (aktualizační balíky, linux distribuce, demo her, apod.) – čím více lidí data stahuje, tím rychleji budou stáhnuta.
- Princip je založen na souběžném stahování malých bloků dat a jejich současné nabízení ke stáhnutí
- Vyhledávání
  - získání metadat (.torrent) z obecného prostředí internetu (WWW, e-mail, FTP, atd.)
  - metadata obsahují informace o souboru (velikost, název, kontrolní součty, atd.) a především URL trackeru
- Stahování
  - komunikace s trackerem a zjištění aktuálních informací o souboru, klientech, kteří jej stahují, atd.
  - realizováno protokolem HTTP (6969)
  - přímá komunikace s „peer“ a „seed“ uzly po blocích, a to obvykle při spolupráci 1:1 (něco za něco), řešení politiky poskytování bloků
- Pojmy
  - tracker – server, který řídí sdílení souborů
  - seed – uzel s kompletní kopií souboru (vždy musí v síti existovat alespoň jeden)
  - peer – uzel, který má jen některé bloky souboru a snaží se o jeho stáhnutí
  - swarm – množina uzlů, které zpracovávají (stahují/posílají) jeden torrent

# BitTorrent

- Problémy

- není zaručena aktuálnost informací v torrentech, vzhledem k jejich „off-line existenci“









# Problémy výměnných P2P sítí

- Technologické a technické
  - rostoucí počet uživatelů, souborů, dotazů
  - zajištění aktuálnosti sdílených informací
  - efektivní algoritmy pro vyhledávání (Google)
  - přístupnost v sítích
  - zabezpečení (přenosu, dat, jednotlivých klientů)
- Etické
  - leeching
  - podvrhování obsahu
  - zahlcování a útoky na sítě
- Právní (MPAA, RIAA, OSA, BSA)
  - samotný princip sítí je legální, problém je ve sdíleném obsahu
  - nejlépe je na tom BitTorrent, protože „peer“ nabízí pouze bloky dat, až „seed“ sdílí kompletní autorské dílo

# Cloudová úložiště

- Dnes velmi rozšířená služba umožňující
  - Ukládání souborů do „cloudu“
  - Sdílení mezi zařízeními
  - Sdílení mezi uživateli
  - Synchronizace a verzování
  - Integrace do aplikací
- Využívá se kombinace různých přístupů a komunikačních protokolů
  - REST over HTTP/HTTPS
  - Delta encoding
  - Vazba na další technologie jako FTP, WebDAV, atd.

	Strengths	Weaknesses	Personal	Business
	<ul style="list-style-type: none"> <li>•Ease of use</li> <li>•Multiple devices</li> <li>•Automatic cloud backup</li> </ul>	Fewer IT admin controls than other cloud storage services	<b>2 GB:</b> free <b>1 TB:</b> \$9.99/mo or \$99/yr	<b>Pro for individuals,</b> 1 TB: \$8.25/user/mo <b>Business for teams,</b> "as much space as needed": \$12.50/user/mo
	Administrative controls	Can seem overwhelming to use for non-technical employees	<b>Personal Pro,</b> 100 GB: \$10/mo	<b>Starter,</b> 100 GB: \$5/user/mo <b>Business,</b> unlimited storage: \$15/user/mo (minimum 3 users)
	<ul style="list-style-type: none"> <li>•Part of Google Apps</li> <li>•Collaboration and sharing</li> </ul>	No option to buy Google Drive as a standalone service for business	<b>15 GB:</b> Free <b>100 GB:</b> \$1.99/mo <b>1 TB:</b> \$9.99/mo <b>10 TB:</b> \$99.99/mo	<b>Google Apps,</b> 30 GB: \$5/user/mo <b>Google Apps</b> with unlimited storage and Vault: \$10/user/mo
	<ul style="list-style-type: none"> <li>•Designed for business</li> <li>•Cloud storage available a la carte or as part of Office 365</li> </ul>	File organization system takes some getting used to	<i>n/a</i>	<b>OneDrive for Business</b> 1 TB: \$5/user/mo 5 TB: \$10/user/mo (minimum 5 users) <b>Office 365 with Cloud Storage</b> Business Essentials, 1 TB: \$5/user/mo (email not included) Business, 1 TB: \$8.25/user/mo (includes email and other apps) Business Premium, 1 TB: \$12.50/user/mo (includes email and all Office 365 apps plus fully installed Office Suite on PC or Mac)
	<ul style="list-style-type: none"> <li>•Flexibility</li> <li>•Options for developers</li> <li>•Pay as you go pricing available</li> </ul>	Complexity makes it difficult to use for non-technical users	<i>n/a</i>	Amazon has a wide variety of pricing plans. You could pay as little as a few pennies a month for the service, or as much as hundreds of dollars per month, depending on how you use the service.
	Works well across multiple Apple devices	Designed for personal use, not business purposes Does not integrate with other services Difficult to share files or collaborate	<b>5 GB:</b> Free <b>50 GB:</b> \$0.99/mo <b>200 GB:</b> \$2.99/mo <b>1 TB:</b> \$9.99/mo	<i>n/a</i>



# Sdílení informací

- Sdílení souborů vs. sdílení informací
  - soubory – fyzické bloky dat nesoucí dané informace
  - informace – informační obsah je přenášený a uložený různými způsoby
- Sdílení informací je požadavek především ze strany organizací
- Nástroje pro podporu sdílení
  - rychlá výměna zpráv (MLM, ICQ, atd.)
  - telefonní a video hovory (MLM, Skype, ICQ, atd.)
  - sdílení souborů pomocí komunikačních kanálů (MLM, ICQ, atd.)
  - sdílení souborů s ohledem na vývoj software (CVS, SVN, GitHub)
  - on-line spolupráce (WebDAV, SharePoint, atd.)
  - Sociální sítě
  - Cloudové služby

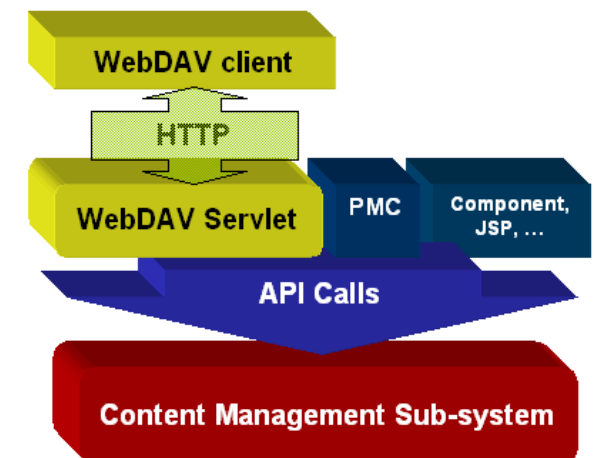
# On-line spolupráce

- Přístup k využití služeb internetu pro podporu spolupráce, koordinace, řízení, apod., a to bez ohledu na geografickou vzdálenost jednotlivých uživatelů
- Obvykle zahrnuje problematiku
  - dokument managementu
  - „time“ managementu
  - projekt management
  - intranetových a extranetových portálů
  - online komunikace
- Principiálně se jedná o takové přístupy, které umožňují kolaborativní práci více uživatelů nad jedněmi daty, a to zpravidla v reálném čase.
- Ve velké míře myšlenky Webu 2.0 umožňují praktickou aplikaci on-line spolupráce
- Dostupná řešení
  - kancelářské aplikace – Google Wokspace, Office365
  - publikační portály – Wiki, CMS, atd.
  - podnikové systémy – SharePoint, WebDAV



# WebDAV

- Web-based Distributed Authoring and Versioning
- IETF standard (RFC 4918)
- Sada rozšíření protokolu HTTP/1.1, umožňující skupinové editování a správu souborů (přímo nepodporuje verzování)
- WebDAV je dnes aplikovatelný napříč širokou škálou platforem a systémů (např. Apache mod\_dav)
- Protože řeší obecný přístup k souborům (CMS), je možné skupinově zpracovávat soubory poskytující nejrůznější obsah a využití (informační stránky, kalendáře, kontakty, projektové dokumenty a procesy, obecné dokumenty, apod.)



# WebDAV

- Podporované metody
  - zamykání (sdílené i exkluzivní zámky)
  - metadata (přiřazování k dokumentům, vyhledávání, atd.)
  - hromadné zpracování souborů - kolekcí
  - namespace
- Rozšiřuje také sadu response kódů
- DeltaV – rozšíření pro verzování
  - rozšíření WebDAV o funkcionality verzování a konfigurace

## **DeltaV** (RFC 3253)

CHECKIN, CHECKOUT, UNCHECKOUT, VERSION-CONTROL, REPORT, UPDATE, LABEL, MERGE, MKWORKSPACE, BASELINE-CONTROL, MKACTIVITY

## **WebDAV** (RFC 2518)

LOCK, UNLOCK, PROPFIND, PROPPATCH, COPY, MOVE, MKCOL

## **HTTP 1.1** (RFC 2616, RFC 2617)

GET, HEAD, POST, PUT, DELETE, OPTIONS, TRACE, CONNECT

# Další koncepty sdílení

- Internet of things – sdílení „mikrodat“
- Cloudová řešení – infrastruktura pro „provoz“ sdílení
- Sdílení výpočetního výkonu – výpočty, blockchain
- Sociální sítě – sdílení příběhů
- „Metaverse“
  
- Sdílení objektů reálného světa – auta, kola, kanceláře, ...