

Corso Di Linux

Liceo Fermi, Bologna - Marcello Galli, novembre 2009

Il web

Oggi l'applicativo piu' diffuso in rete e' il web (www); e' cosi' diffuso che molta gente crede che il web sia internet, e che l'unico modo di usare internet sia tramite Internet Explorer della Microsoft, o, al piu', Firefox. Invece questo e' solo uno dei tanti applicativi che lavorano in rete: e' un sistema di distribuzione di contenuti vari (testi, suoni, immagini, video) basata su un sistema client-server.

Ci sono quindi 2 componenti del sistema:

- il server web: un programma che, a richiesta, fornisce dati, organizzati in pagine, utilizzando il protocollo http. Il piu' diffuso programma server e' "Apache", che si trova in tutte le distribuzioni Linux. La Microsoft ha IIS (Internet Information Services), ben noto per innumerevoli problemi di sicurezza.

Il protocollo di comunicazione usato per trasmettere le pagine si chiama http, ed e' un protocollo molto semplice. Questo e' uno dei motivi del suo successo.

- il client web, detto anche browser, che chiede le pagine al server e quando gli arrivano le visualizza sullo schermo. Ci sono tanti diversi programmi client, giusto per citarne alcuni: mozilla, firefox, konqueror, lynx, opera, epiphany, internet explorer etc.

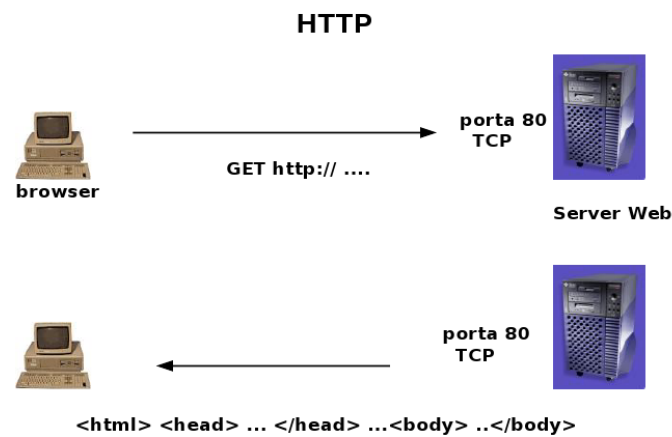


Figure 1: Comunicazione con protocollo http

La peculiarita' delle pagine web sono i links: le pagine possono contenere indirizzi di altre pagine, ed il browser le chiede al server ove sono quando uno clicca sul link. In questo modo si puo' "navigare" passando da una pagina all'altra senza un particolare ordine e senza che pagine debbano avere una

struttura prefissata, come ad esempio una struttura gerarchica (era così in sistemi più vecchi, come gopher). Per questa organizzazione delle pagine, legate da links, si dice anche che il web è un "ipertesto".

Questo modo di distribuire informazioni è molto comodo, per cui si è diffuso moltissimo ed i client sono diventati capaci di fare da interfacce anche ad altri programmi, come i server ftp, i server di posta etc. Alla fine diventano strumenti unificati a tutti i servizi della rete. Il protocollo http facilita questo, fornendo una sintassi unificata per le risorse di rete, che si chiama URL (uniform resource locator).

Una pagina web ha un indirizzo tipo:

<http://www.helldragon.org/index.html>

La specifica completa sarebbe:

protocollo://username:password@host:porta/directory/file.ext?parameter=value&par=value#anchor

i campi hanno il seguente significato:

- protocollo: un protocollo è un insieme di regole per la comunicazione; il genere trovate http, che indica come comunicano un client ed un server web, ma potete trovare anche ftp, ed allora il vostro client sta facendo un trasferimento con ftp, o mail, per mandare posta, od anche altre cose; il protocollo è seguito dai caratteri '://'
- username:password: certe pagine possono richiedere un username ed una password per l'accesso, vanno messe dopo il protocollo;
- host: il nome o l'indirizzo IP del computer su cui corre il programma server, nell'esempio sopra è: www.helldragon.org. Molte macchine che hanno un server web hanno nome www, ma non è una regola, potete mettere un server su una macchina con un nome qualunque;
- porta: la porta TCP su cui è in ascolto il server, per convenzione è la porta 80, ma è chi configura il server che decide la porta e può mettere la porta che vuole; se non si mette nulla il client chiede alla porta 80;
- directory: la directory ove si trova il file, è il percorso completo della directory, partendo dalla cartella ove sono le pagine web da distribuire; se ad esempio le pagine web sono in /var/www e cercate il file /var/www/ddd/index.html dovete indicare: ddd/index.html;
- file.ext: il nome del file richiesto; non è una regola, ma l'estensione è in genere html od htm, per pagine web, txt per semplici testi, php od asp quando interpellate programmi sul server; in quest'ultimo caso il file richiesto contiene un programma che corre sul server e produce una pagina web, che viene mandata al client;
- ?parameter=value&par=value: in caso si acceda ad un programma invece che ad una semplice pagina html, si possono fornire con questa sintassi parametri al programma; il programma crea una pagina, che viene mandata al client. Questo modo di interfacciarsi con programmi che corrono sul server è detto: CGI (Common Gateway Interface);
- #anchor: indica una posizione entro il file; il browser non si posizionerà ad inizio pagina, come di solito, ma nella posizione specificata.

Le pagine web non sono difficili da fare, sono normali testi, con in mezzo indicazioni particolari (tag) che dicono come la pagina deve essere mostrata, ad esempio abbiamo i caratteri: <H1> che indicano un titolo grande; <H2> per un titolo piccolo, <P> per un paragrafo,
 per andare a capo, per mettere una figura e così via. Pagine fatte in questo modo si dice che sono scritte in linguaggio HTML (Hyper Text Markup Language). I Browser hanno un'opzione che vi fa vedere la pagina HTML come è veramente, e come il server la manda al client.

All'inizio i siti web erano fatti di pagine semplici, poi le esigenze sono aumentate e sono stati introdotte specifiche di formattazione più dettagliate (i CSS: cascade style sheet), per chi ha esigenze editoriali.

Poi si volevano effetti ed animazioni nella pagina e si voleva che certe operazioni, come controlli su dati da passare al server, le facesse il client, risparmiando lavoro al server. Allora è stato introdotto javascript; un linguaggio di programmazione con istruzioni inserite nella pagina web; quando il client le trova le esegue. In questo modo si fanno menu eleganti, effetti sulle figure ed altre cose simili.

Poi i server sono diventati piu' veloci e sono nate tecniche per sfruttare il lavoro del server; sul server ci sono programmi che scrivono le pagine quando vengono richieste dal client, prendendo i contenuti da un database; molti siti oggi sono fatti in questo modo, ad esempio quelli dei giornali. La tendenza corrente per la tecnologia web e' quindi di avere:

- un computer su cui corre un server web, tipo Apache;
- un database (archivio di dati con indici, organizzati in tabelle collegate),
- un programma abbinato al server, che, usando i dati del database, fa le pagine al volo quando vengono richieste.
- le pagine create possono contenere istruzioni javascript, perche' il client le esegua;
- un client, che richiede le pagine al server ed esegue le istruzioni javascript che ci trova dentro;

Sono tecnologie abbastanza complicate, ma per fare pagine semplici non servono, basta usare la sintassi del linguaggio HTML.

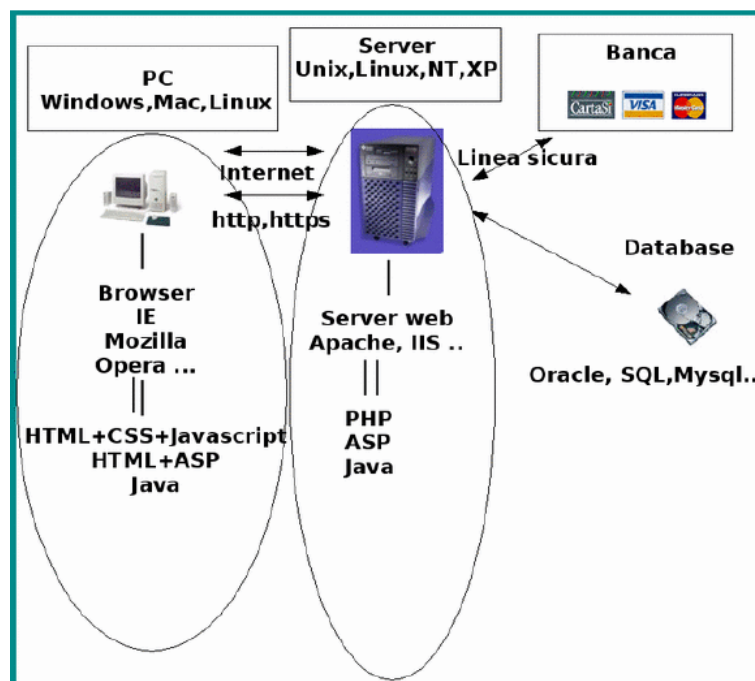


Figure 2: Tecnologie web

Il Mail

Un'altra applicazione molto usata e' il mail. Anche questa viene spesso usata tramite un'interfaccia web, cioe' tramite il browser e potete accedere alla vostra posta col browser, oppure scaricarla sul vostro computer.

Il mail e' basato su un modello client-server. Ci sono dei programmi server che ricevono la posta, la smistano e la mandano a destinazione, e dei programmi client, che leggono la posta.

POP ed IMAP sono programmi che gestiscono caselle di posta su un server, un programma client puo' recuperare la posta che si trova su questi server od organizzarla in cartelle sul server stesso (con IMAP).

Ci sono tanti programmi client per la posta: Eudora, Thunderbird, kmail etc. etc., che mandano la posta ad un server POP od IMAP, oppure la recuperano dai server e la mostrano all'utente, o la conservano in cartelle apposite. Il protocollo con cui si manda la posta al server si chiama SMTP (simple mail transfer protocol), ed usa la porta 25.

Altre applicazioni che usano tecnologie web

Le tecnologie legate al web sono diventate un modo comune di utilizzare le applicazioni di rete e le applicazioni di rete sono tantissime, infatti la rete e' sempre piu' usata ed il computer diviene un "terminale" utilizzato in un sistema di interconnessioni globale.

Ci sono programmi di rete per gestire archivi, sincronizzare cartelle, fare backup, connessioni criptate, chat e tantissimi programmi commerciali, per negozi, banche etc. etc., sembra che fra un po' la maggior parte del software usera' la rete e molto di questo software si utilizza tramite un browser web.

Applicazioni di rete molto di moda oggi sono i sistemi peer-to-peer, ove si scambiano dati senza che questi debbano risiedere su un server centrale, da cui tutti li prendono.

Importante e' anche l'uso di internet per telefonia. I segnali telefonici sono codificati in sequenze di bit e trasmessi via Internet (si parla di VoIP: Voice over IP). Spesso questi sistemi permettono di risparmiare sui costi telefonici; ci sono anche veri e propri sistemi di videoconferenze che utilizzano internet, come skype.

L'ultima novita' sono i social networks, come facebook o linkedlin, non sono altro che applicazioni web, con un server web, ed un database dietro, fatte per facilitare i contatti fra le persone.

© Marcello Galli, Novembre 2009. Sito di riferimento: <http://www.helldragon.eu>