

# Siti web centrati sui dati (Data-centric web applications)

1

ALBERTO BELUSSI  
ANNO ACCADEMICO 2012/2013

# WEB

2

La tecnologia del World Wide Web (WWW) costituisce attualmente lo strumento di riferimento per la diffusione e acquisizione di informazioni sotto forma di documenti digitali.

Tale tecnologia si basa sul sistema di comunicazione planetario chiamato INTERNET.

# INTERNET

3

E' una federazione di reti che comunicano attraverso la famiglia di protocolli TCP/IP.

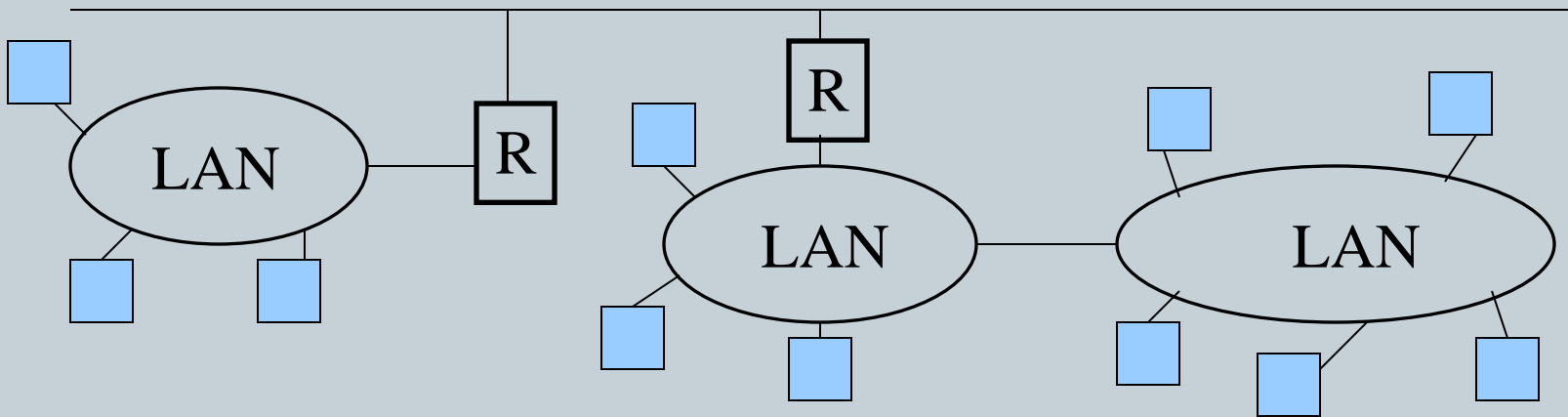
Riferimenti storici:

- ARPANET (dipartimento della difesa USA): è stata la prima rete a commutazione di pacchetto basata sul protocollo IP
- NSFnet (national science foundation net)

# Struttura attuale della rete

4

Internet è attualmente un insieme di reti locali (Local Area Network) collegate tra loro e/o con altre porzioni di rete attraverso ROUTER (calcolatori dedicati alla gestione della rete) o dispositivi simili.



# Struttura attuale della rete

5

I nodi della rete sono i calcolatori connessi ad una delle reti federate.

Ogni calcolatore ha un indirizzo univoco, detto indirizzo IP, che ha la seguente forma:

157.27.252.11

Tale indirizzo può assumere anche un formato simbolico:

dbserver.scienze.univr.it

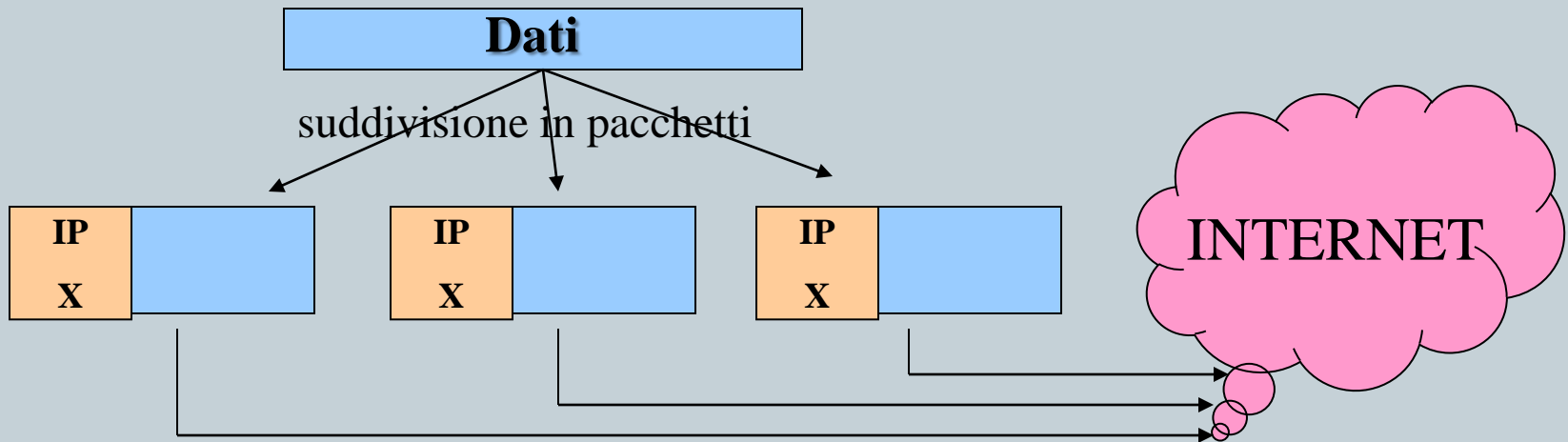
# Struttura attuale della rete

6

Come avviene la comunicazione sulla rete?

Meccanismo di base:

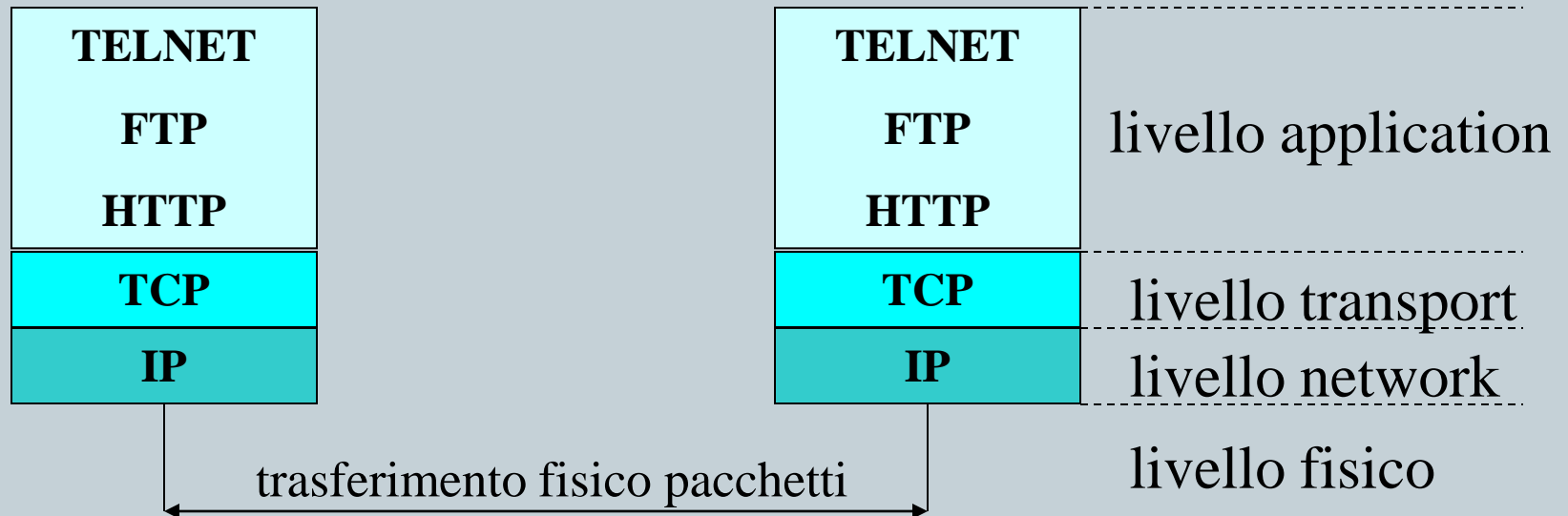
- commutazione di pacchetto
- scambio di informazione organizzata in pacchetti



# Struttura attuale della rete

7

Livelli del Software di Rete secondo il protocollo TCP/IP.



# Struttura attuale della rete

## Protocollo TCP/IP:

- Il protocollo IP (Internet Protocol) garantisce la spedizione del pacchetto a destinazione
- Il protocollo TCP (Transmission Control Protocol) gestisce la suddivisione in pacchetti dell'informazione da spedire, garantisce la spedizione senza errori e gestisce il riassettaggio dei pacchetti nell'ordine corretto.
- I protocolli del livello application gestiscono l'interazione via rete tra due calcolatori (nodi della rete). Si basano sul paradigma CLIENT-SERVER.



# Protocollo HTTP

9

Il protocollo HTTP gestisce i servizi collegati al WEB. L'informazione disponibile in rete attraverso il protocollo HTTP è organizzata in documenti digitali detti IPERTESTI.

HTTP significa infatti:

**HyperTEXT TRANSFER PROTOCOL.**

# IPERTESTI

10

**IPERTESTO:** è un documento con struttura non sequenziale, costituito da varie parti fra loro collegate. Tali legami sono detti LINK e consentono di navigare nell'ipertesto.

Le singole parti di un ipertesto si dicono **PAGINE WEB**.

L'insieme di pagine web gestite da un unico server HTTP della rete costituiscono un **SITO WEB**.

# Linguaggio HTML

11

Il linguaggio usato per specificare IPERTESTI si chiama: HTML (HYPERTEXT MARKUP LANGUAGE).

Ogni pagina web diventa un file HTML; l'HTML consente di specificare:

- La formattazione del testo per la presentazione
- La struttura della pagina (sezioni, tabelle, liste, ecc.)
- I legami con altre pagine (LINK) o con sottoparti della pagina
- Il contenuto informativo della pagina.

# Linguaggio HTML

12

## LINK in HTML

Come è possibile identificare in modo univoco un documento (risorsa) nel World Wide Web?

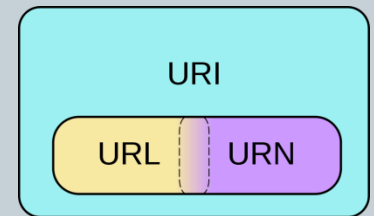
Si introduce il concetto di URL (Uniform Resource Locator)

*protocollo://server/risorsa*

*protocollo:* ftp, mailto, http

*server:* ip o nome simbolico del nodo

*risorsa:* [path]nome\_file[parametri]



# Linguaggio HTML

13

## LINK in HTML:

- Link **esterno**:

```
<a href="URL"> testo </a>
```

- Link **interno**:

```
<a href="URL#label"> testo </a>
```

...

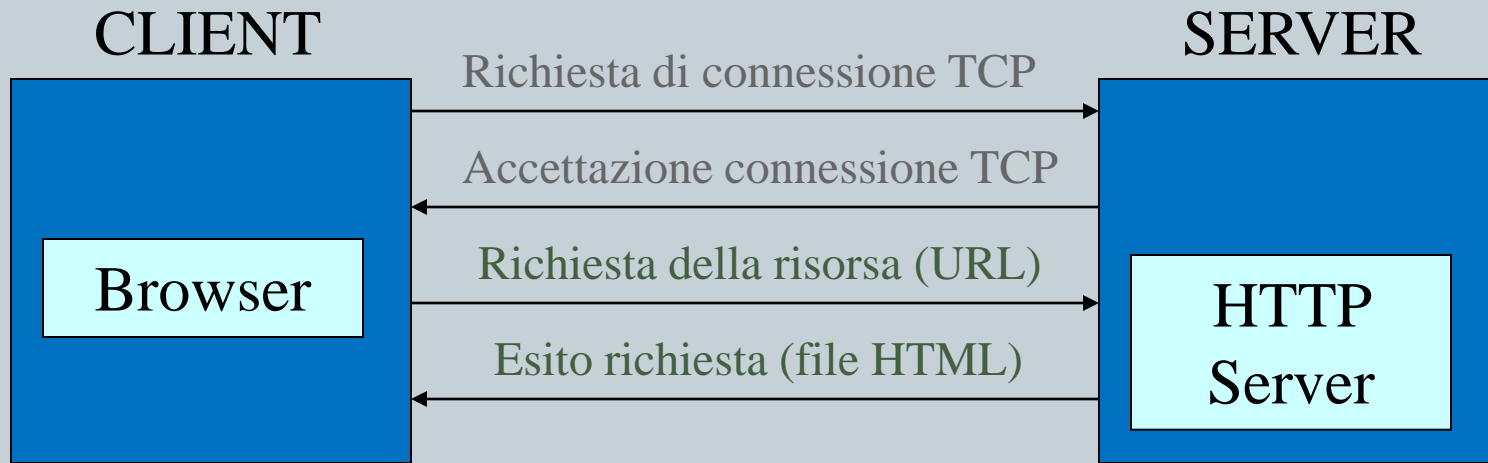
...

```
<a name="label"> testo 2 </a>
```

# Protocollo HTTP

14

Consente lo scambio tra client (browser) e server (HTTP server) di pagine web (file HTML).



# Protocollo HTTP

15

Il protocollo HTTP è “state-less” (senza stato): ogni connessione è completamente indipendente dalle altre. Il server non è in grado di mantenere informazioni sulle connessioni passate.

## Richieste HTTP:

- GET: richiede una pagina (URL) inviando i parametri in modo esplicito in coda all'URL.
- POST: richiede un pagina (URL) con invio dei parametri in modo non visibile nell'URL.
- PUT, HEAD, DELETE, TRACE, OPTIONS.

# Gestione delle pagine WEB

16

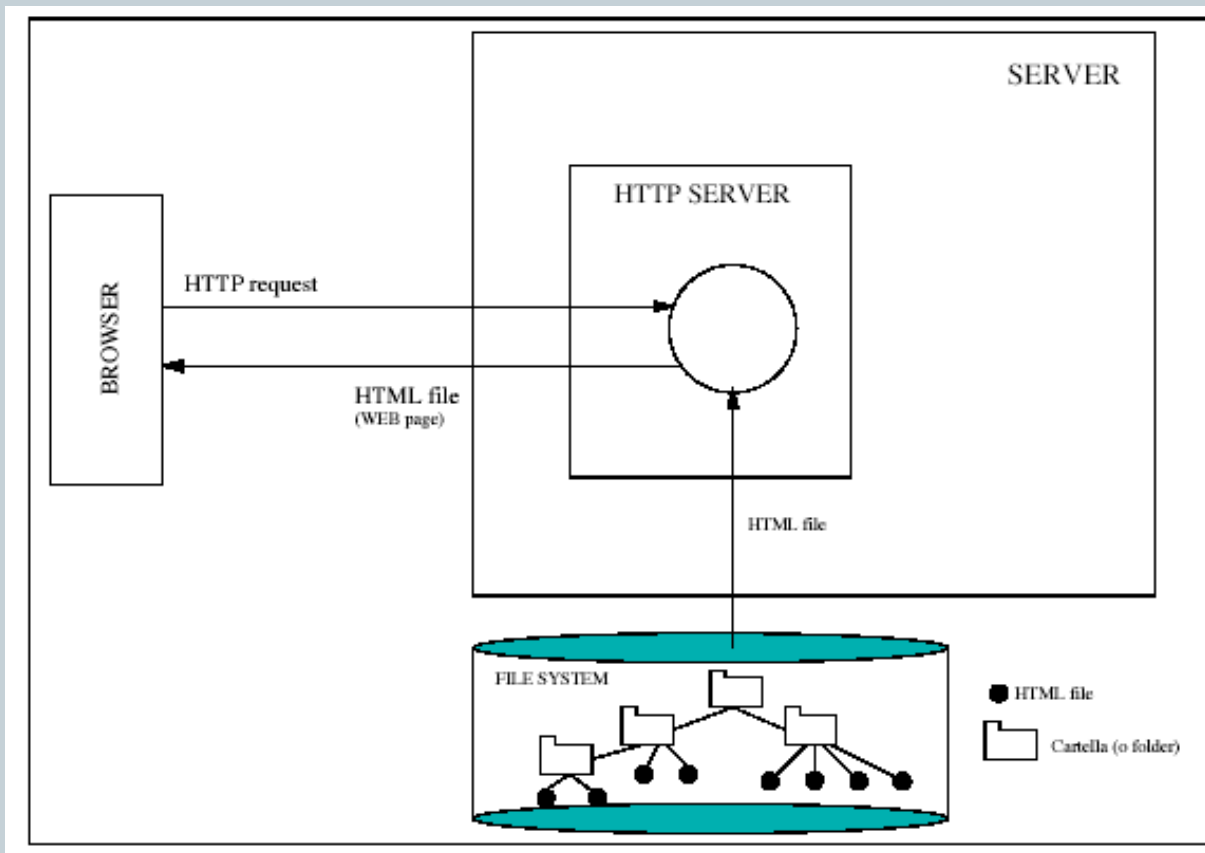
- **Pagine statiche**
  - Sono File HTML memorizzati nel file system del server
- **Pagine dinamiche:**
  - Vengono generate “on the fly” attraverso l’esecuzione di programmi sul server;
  - Tali programmi spesso interagiscono con un DBMS che contiene i dati da mostrare nelle pagine.



# Pagine statiche

17

In questo caso le pagine vengono generate da un programmatore HTML e vengono rese disponibili sul server HTTP come file HTML.



# Pagine dinamiche

18

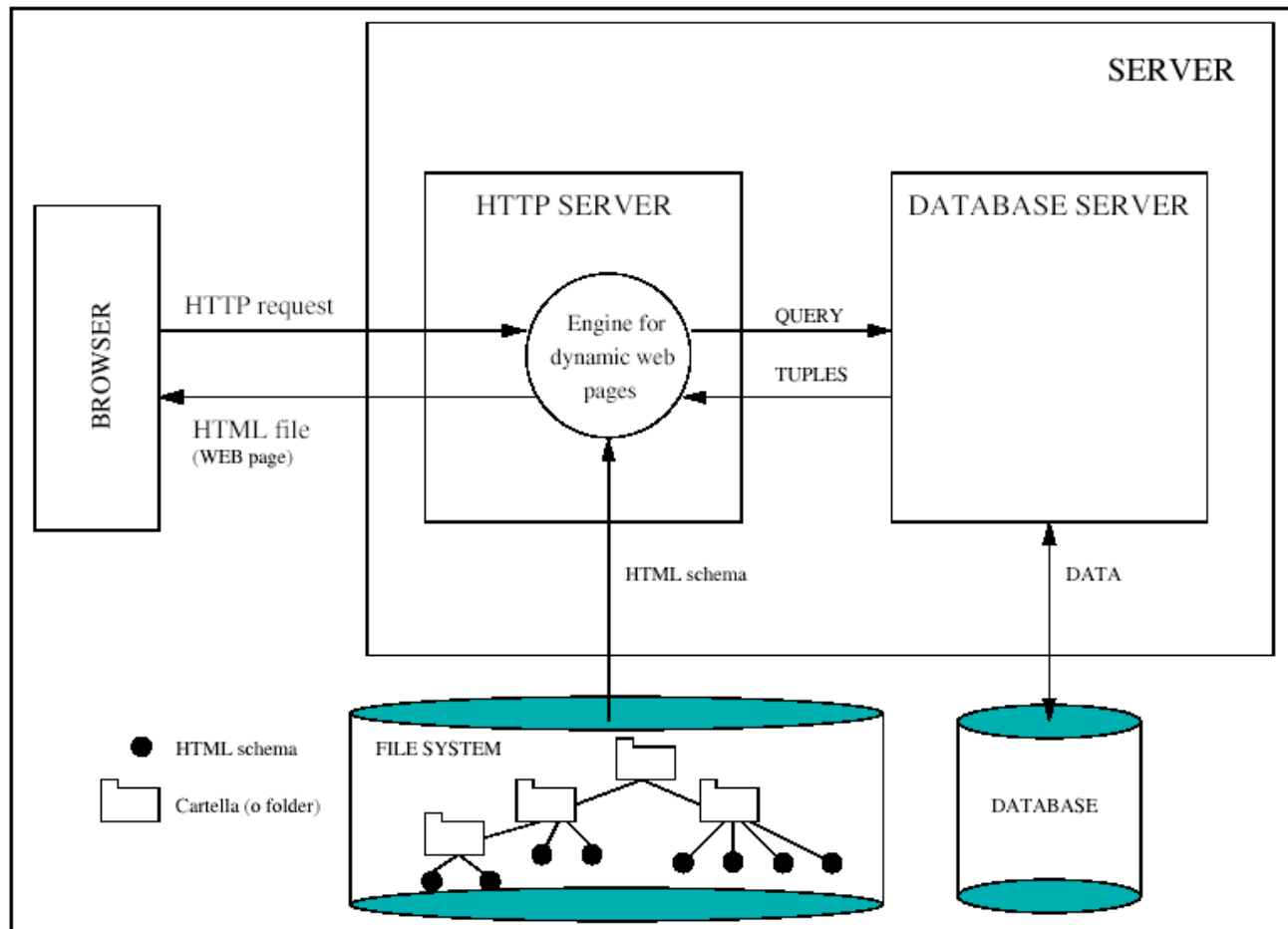
In questo caso invece le pagine vengono generate dal server in base ad altre applicazioni esterne o integrate nel server HTTP.

La generazione dinamica delle pagine consente di estrarre da sistemi DBMS l'informazione da presentare.

Tali DBMS raccolgono gli aggiornamenti dei dati prodotti dai processi "core" dell'organizzazione per cui si sta allestendo il sito web.

# Pagine dinamiche: schema

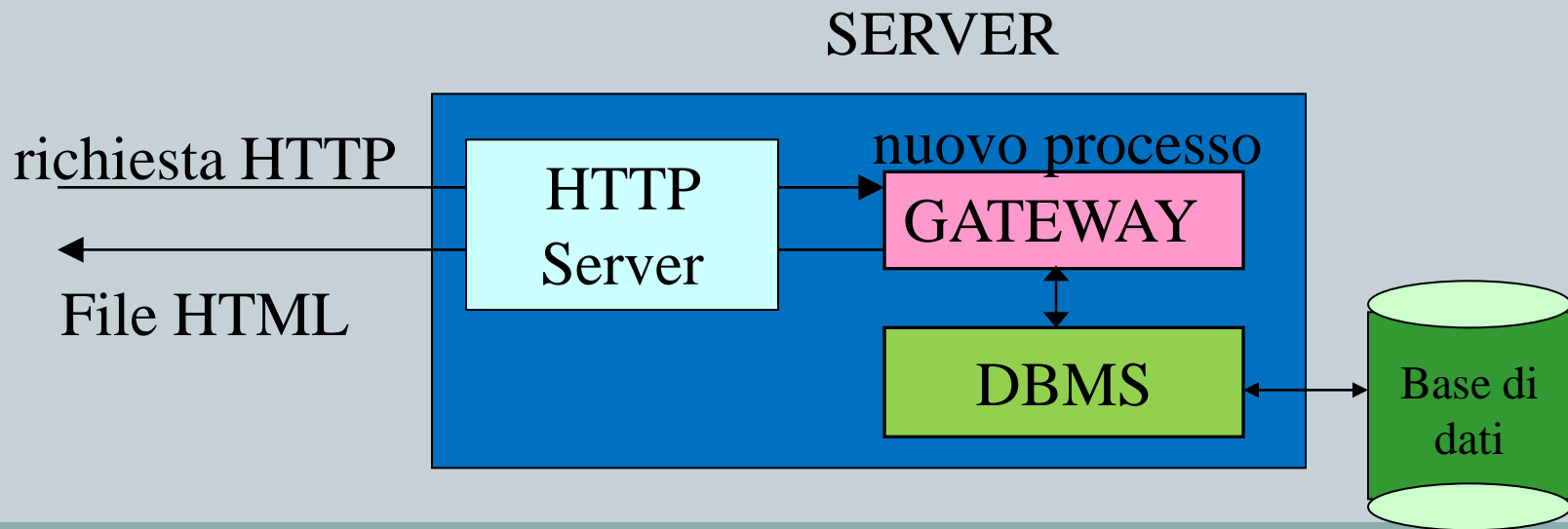
19



# Tecniche per realizzare pagine web dinamiche: CGI

20

Esistono attualmente diverse tecniche per realizzare pagine web dinamiche, ma tutte derivano dalla prima tecnica applicata storicamente: CGI (Common Gateway Interface). Tale tecnica utilizza l'URL della richiesta HTTP per invocare un programma presente sul server, che calcolerà la pagina HTML da restituire al client.



# Tecniche per realizzare pagine web dinamiche CGI

## LIMITI DELL'APPROCCIO CGI

- Richiede la creazione di un **nuovo processo per ogni richiesta**: tale creazione richiede tempo non trascurabile e necessita inoltre di uno specifico spazio di indirizzamento in memoria virtuale.
- Il processo CGI attivato richiede necessariamente una nuova connessione al DBMS e la sua corrispondente chiusura a fine programma.
- Non è possibile condividere dati tra richieste HTTP successive.

# Tecniche per realizzare pagine web dinamiche: CGI

22

## Tecniche alternative a CGI

- Soluzioni basate sull'estensione del server HTTP che diventa un **application server**: questo ad esempio è il caso della tecnologia **Servlet di Sun (ora Oracle)**.
- Soluzioni basate sull'immersione di codice nei file HTML: **Java Server Pages (JSP)**, Active Server Pages (ASP), **Hypertext Preprocessor (PHP)**.

# Tecniche per realizzare pagine web dinamiche: SERVLET

23

**SERVLET:** si basa sul linguaggio JAVA.

Vantaggi rispetto a CGI:

- Ogni richiesta genera un thread e non un processo.
- È possibile mantenere connessioni aperte con il DBMS tra una richiesta e l'altra.
- Portabilità JAVA.

# Tecniche per realizzare pagine web dinamiche: TEMPLATE SYSTEMS

24

**Template Systems:** si basano su versioni di HTML estese con marcatori (tag) per l'inserimento di codice.

**Vantaggi rispetto a CGI:**

- Tutti i vantaggi delle servlet.
- Il codice per la generazione delle parti dinamiche viene inserito con opportuni tag nel file HTML che rappresenta invece la parte statica della pagina WEB. Evita al programmatore di generare dinamicamente anche le parti fisse.
- Rientrano in questa categoria: **JSP** (Java Server Pages), **PHP** (Hypertext Preprocessor), ASP, ecc...