Le Web et les bases de donées

C. Delobel, V. Benzaken, E. Waller

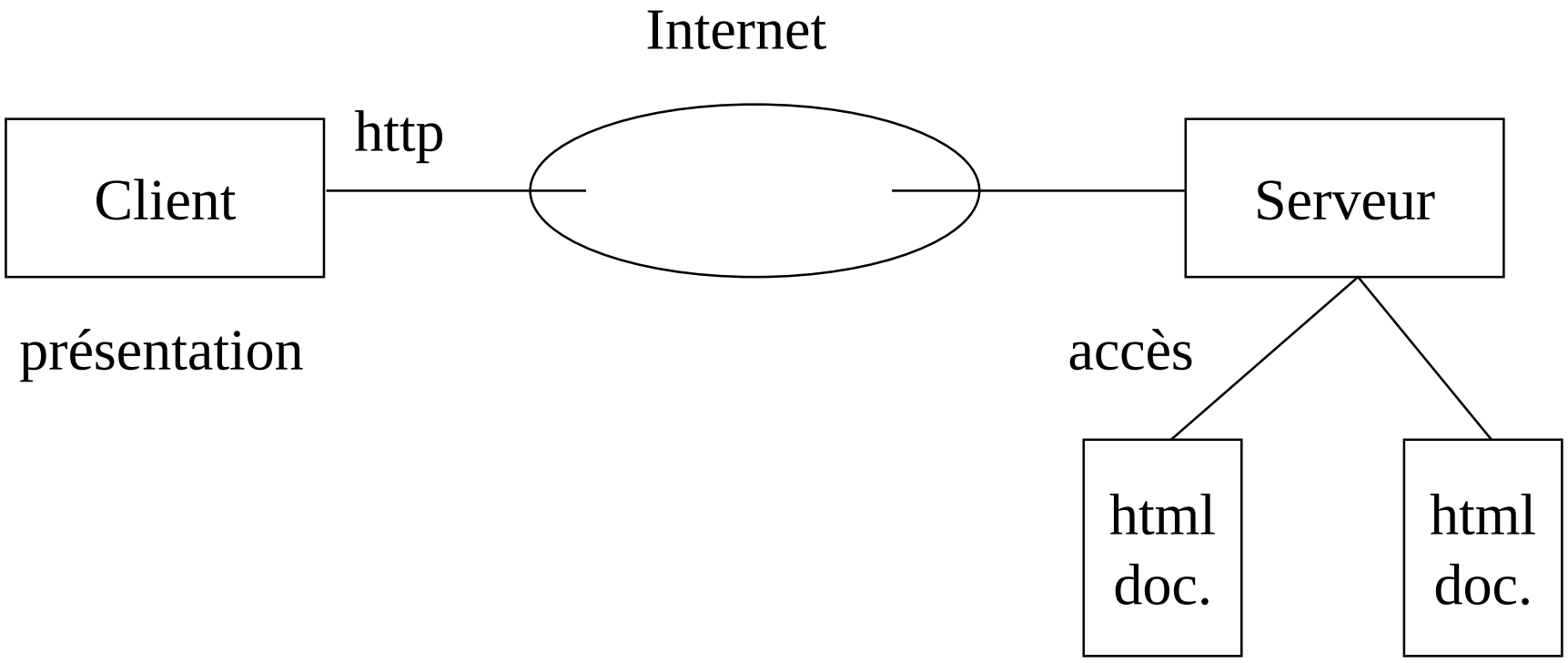
Le point de départ

- Web
  - outil d'échange et d'accès à l'information (http)
  - l'unité d'information est la page (fichier)
  - possibilité de navigation au travers des adresses (URL)
- Base de données
  - modèle et schéma
  - langage de requêtes
  - transaction et reprise

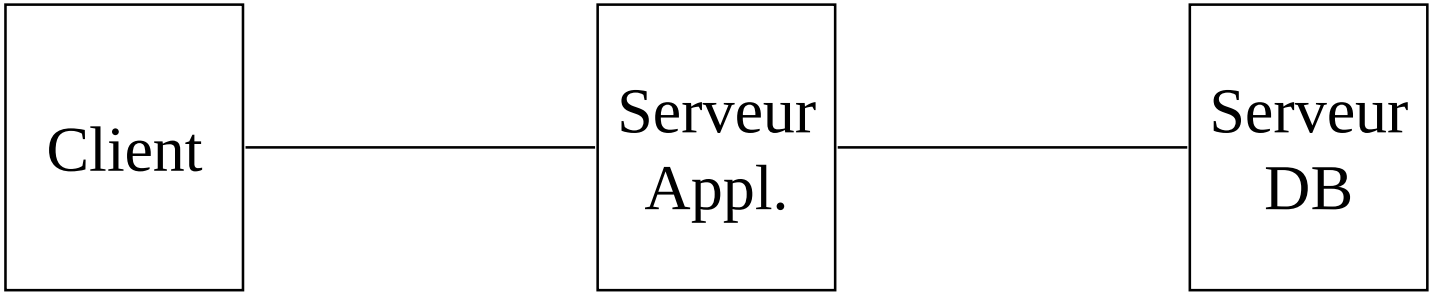
Une intégration : pourquoi ?

- Les données sont dans des sources diversifiées (base de données, fichiers, tableurs, etc.).
- Le transfert d'une source à une autre est toujours une tâche problématique.
- HTML est plus un standard de présentation des données que de structure et d'échange.
- Un format d'échange et de structuration faible des données comme XML est une orientation nouvelle.

Architecture deux niveaux



Architecture trois niveaux



présentation

application

accès et validation

Protocole sans état

- HTTP est un protocole fondé sur le principe question-réponse :
  - ➤ établissement d'une connexion avec le serveur web
  - ➤ le client envoie une requête
  - ➤ le serveur répond à la requête sous forme d'une page html
  - ➤ le serveur clos la connexion
- Le concept de transaction permettant d’englober plusieurs questions-réponses dans la même unité n'existe pas.

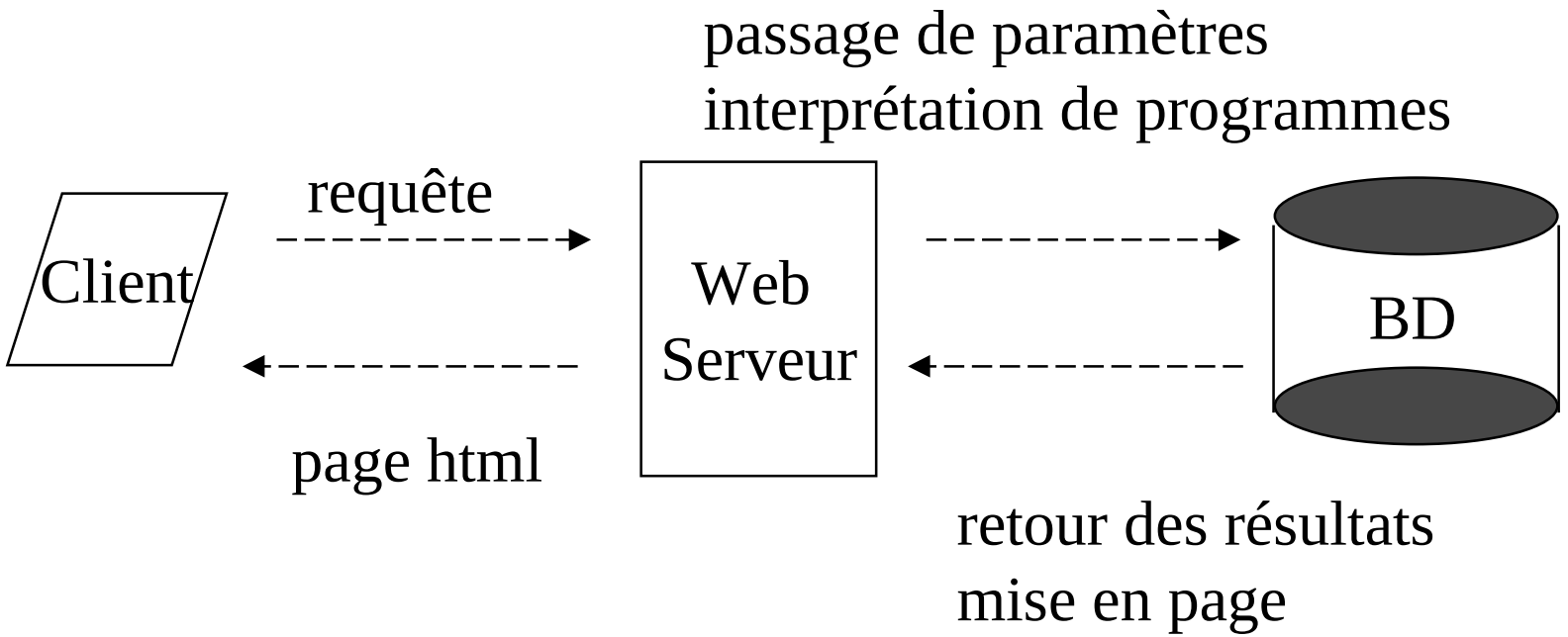
Les besoins

- Accès sécurisé aux bases de données avec des mécanismes d'authendification.
- La solution doit être indépendante de tout choix de systèmes de bases de données, des outils de connectiques et du feuilleteur Web.
- Fournir un mécanisme transactionnel.
- Des outils de développement des applications.
- L'utilisateur doit pouvoir effectuer des recherches d'information.

Les contraintes actuelles

- La sécurité est une préoccupation essentielle.
- La fiabilité : on ne garantit pas le retour d'une réponse.
- Les performances : pointe de connexion et trop de mécanismes d'interprétation.
- Coût de développement : on estime à plusieurs million de francs le coût de maintenance d'un site Web.
- HTML est trop limité.

La problématique



HTML : rappels

- HTML: Hypertext Markup Language
- Implémentation de SGML
- Décrit plus la présentation physique d'un document que sa structure logique.
- Langage de marques avec début et fin : <TAG> ..... </TAG>

Page HTML

- <HTML> .... </HTML> limite d'un document
- <HEADER> .... </HEADER> portion n'apparaissant pas à l'écran mais pouvant être exploitée
- <TITLE> .... </TITLE> portion apparaissant en haut de la fenêtre du feuilleteur
- <BODY> .... </BODY> corps du document pouvant être subdivisé en 6 niveaux <H1> à <H6>, sections, sous-sections, ...

Exemple de page HTML

```
<HTML>
<HEAD>
        <TITLE> document1
        </TITLE>
</HEAD>
<BODY>
  <H1> Le titre du document
  </H1>
  <H2> Premier chapitre </H2>
  <H3> Premeière section </H3>
</BODY>
</HTML>
```

Autres marqueurs

- Séparateur de paragraphes <P> .... </P>
- Saut de ligne
- Ligne horizontale <HR>
- La mise en forme des caractères
  - <B> ... </B>, <I> ... </I>, <TT> ... </TT>
  - mettre un texte en valeur
  - mettre en valeur avec une visibilité supérieure
  - etc.

Les listes

- Liste non numérotées <UL> .... </UL>chaque item est précédé du marqueur <LI>
- Liste numérotées <OL> .... </OL>, chaque item est précédé du marqueur <LI>
- Liste de définitions <DL> .... </DL> , les termes sont pécédés de <DT> et les définitions de <DD>

Exemple de liste

```
<UL>
<LI> Les villes françaises
 <OL>
   <LI> Paris
   <LI> Nice
 </OL>
<LI> Les villes italiennes
 <DL>
   <DT> Villes anciennes
   <DD>  Elles sont ...
   <DT> Villes modernes
   <DD> Elles sont ...
 </DL>  </UL>
```

- **Les villes françaises**
  1. **Paris**
  2. **Nice**
- **Les villes italiennes**

  Villes anciennes

  Elles sont ...

  Villes modernes

  Elles sont ...

Les liens hypertexte

- HTML offre la possibilité d'effectuer des liens entre régions d'un document vers un autre document.
- Différents type de liens (URL) :
  - liens relatifs ou absolus
  - liens vers de régions spécifiques
  - liens entre documents ou internes à un document
  - liens réseaux

Liens inter-documents

- Liens relatifs : la localisation se fait par un chemin Unix relatif par rapport au répertoire courant Voulez-vous la <A HREF = "../doc2.html" suite </A> du document
- Voulez-vous la suite du document
- Liens absolus : la localisation se fait par un chemin absolu ; Voulez-vous la <A HREF = "/home/doc/doc2.html" suite </A> du document

Liens réseaux

<protocol>://<host> [:<port>]/<path> [? <parameter>]

- protocol est le protocole de communication : http, ftp, gopher, ...
- host est l’adresse internet du serveur
- path est le chemin d'accès à un document ou à un programme avec ses paramètres

Document comme formulaire

- Il existe un besoin de pouvoir saisir, au niveau du client des paramètres ou des actions, et de les passer au serveur.
- C'est le concept de formulaire qui permet de le faire.
- <FORM METHOD="<methode>" ACTION="<action>" > ..... </FORM>
- Possibilité de définir les champs du formulaire à l'aide des marqueurs <INPUT> ou <SELECT>

Exemple de formulaire

```
.........
<FORM METHOD="GET" ACTION= "http/www.serveur.fr/cgi/quote.pl">
Name:<INPUT TYPE="text" NAME="par1" SIZE=15> <BR>
Password: <INPUT TYPE= "password" NAME="par2" SIZE=6> <HR>
<INPUT TYPE= "submit" VALUE="SUBMIT">
<INPUT TYPE="reset" VALUE="CLEAR>
</FORM>
.......
```

HOME PAGE

Name: Dupont

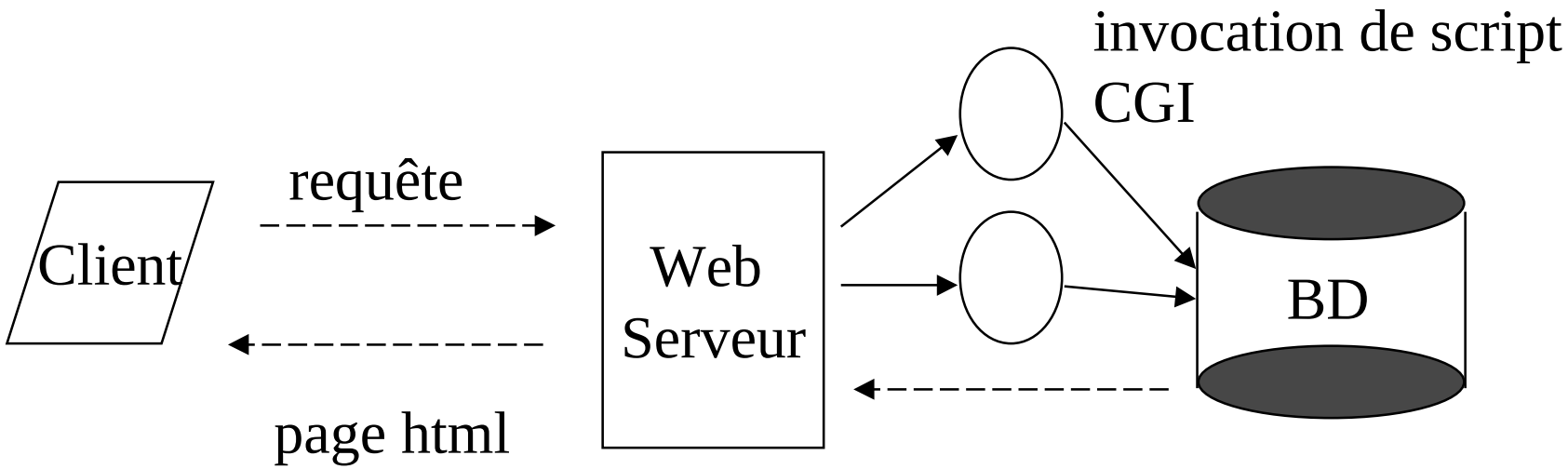
Password: ***

SUBMIT CLEAR

Exemple (suite)

- Lorsque l'utilisateur appuiera sur le bouton SUBMIT après avoir enregistré son nom et son mot de passe on passera au serveur l'adresse suivante :
- http://www.serveur.fr/cgi/quote.pl?par1=Dupont&par2=TMC
- La chaîne passée contient une concaténation des paramètres spécifiés dans le formulaire, cette chaîne peut être interprétée par le serveur par le mécanisme de script CGI.

CGI



retour des résultats et
mise en forme d'une page
html

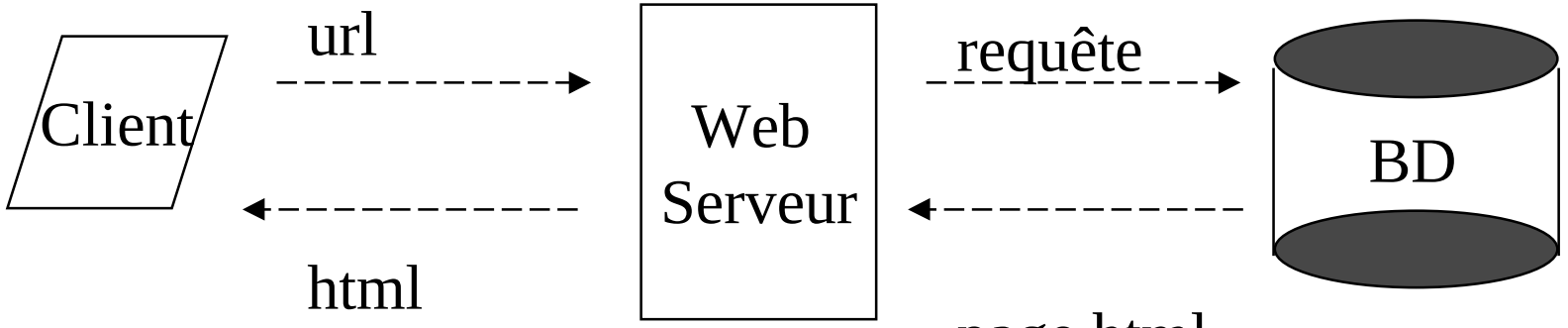
CGI ?

- CGI : Common Interface Gateway
- L'interface définit comment comment les scripts doivent communiquer avec le serveur.
- Un script CGI peut être écrit dans la plupart des langages de programmation, (Perl, C,.. pour Unix).
- Exécution d'un script par le serveur : (1) préparation de variables d'environnments, (2) lecture des variables d'entrée, (3) exécution, (4) retour des résultats.

Limitations

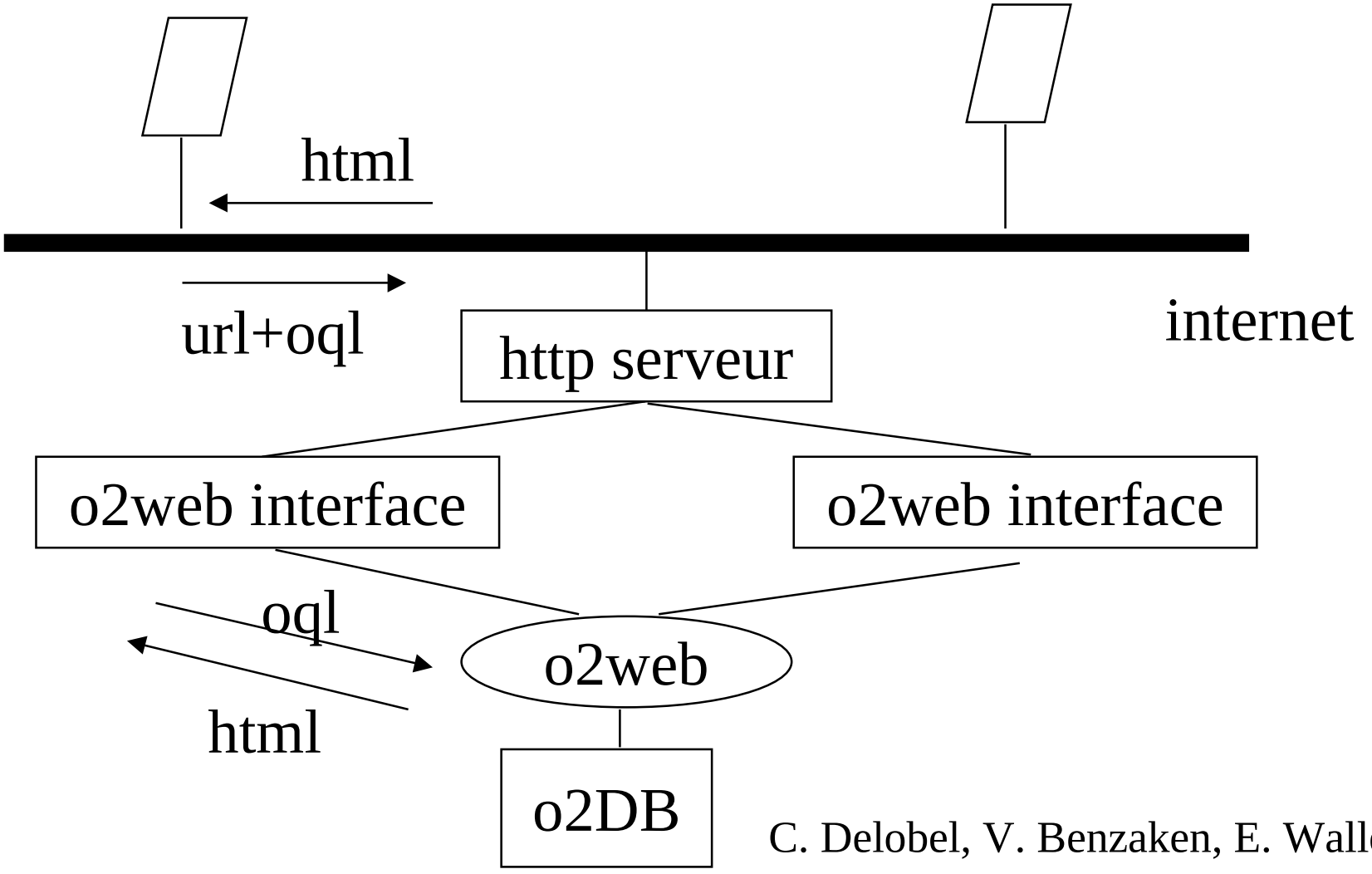
- L’invocation de chaque script CGI est une opération coûteuse car il faut ouvrir et fermer la base de données.
- On peut remédier à cette situation en créant des processus qui laissent la base de données ouverte. L'invocation d'un nouveau script recherche un processus libre.
- Le système est toujours sans état.

Mode requête



page html
dynamique
générée par la BD

O2 : architecture web



Séquence des actions

- 1. Le serveur Web déclenche le script o2webgateway
- 2. L'interface demande à o2web de se connecter à la base de données info de sys et lui passe la requête
- 3. La base de données interprète la requête et passe les résultats à o2web
- 4. o2web convertit dynamiquement les données o2 au format html et les retourne vers le serveur.

http://www.o2tech.fr/o2web/o2webgateway/sys/
info?selct p from p in Lesprojets where p.resume like "www"

Variété des approches

- Incorporer SQL dans HTML
- Connexion entre Java et les bases de données
- Tout les grands développeurs de logiciels ont leur solution : Informix, Oracle, Microsoft, IBM, ObjectDesign, etc.