

Technologies Web

Internet : introduction

Simon CHOLLET

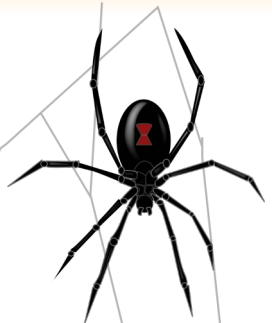


- Définitions
- Histoire
- Évolutions
- Fonctionnement
- Réseau
- Serveurs Web
- Hébergements
- Serveurs particuliers
- Structure et sécurité
- Conclusion



Introduction

- Définitions
- Histoire



Définitions : HTTP, HTTP(S), URL

- **WWW** = **W**orld **W**ide **W**eb :
 - Pour : « la toile (d'araignée) mondiale »
 - Communément appelé : « **WEB** » ou « **la Toile** »
 - Système permettant d'accéder aux **ressources publiques** présentes sur **Internet** et de naviguer à partir de **liens**
 - Liens sont dits « **hypertextes** » → **lier** les pages entre elles
- **HTTP** = **H**yper**T**ext **T**ransfert **P**rotocol
- **HTTPS** = **H**yper**T**ext **T**ransfert **P**rotocol **S**ecured :
 - Protocoles de **communication** entre serveur et client
 - '**S**' pour « sécurisé » → utilisant du chiffrement de données
- **URL** = **U**niform **R**essource **L**ocator :
 - Localiser d'une manière **unique** et **normée** une ressource
 - C'est l'**adresse** que l'on saisit dans son navigateur





Définitions : Internet, Intranet, Extranet

- **Internet** :

- Réseau informatique **mondial** : infrastructure globale, basée sur le protocole **IP** (Internet **P**rotocol)
- D'autres services (autres que le Web) s'appuient sur ce protocole



- **Intranet** :

- Réseau informatique **local** : infrastructure basée aussi sur le protocole IP, mais avec des adresses spécifiques
- Ce réseau n'est **pas visible** de l'**extérieur**



- **Extranet** :

- Utilisation du réseau Internet (publique) pour **interconnecter** plusieurs réseaux locaux (privés)
- Permet l'utilisation de **ressources internes** d'un réseau depuis l'**extérieur**



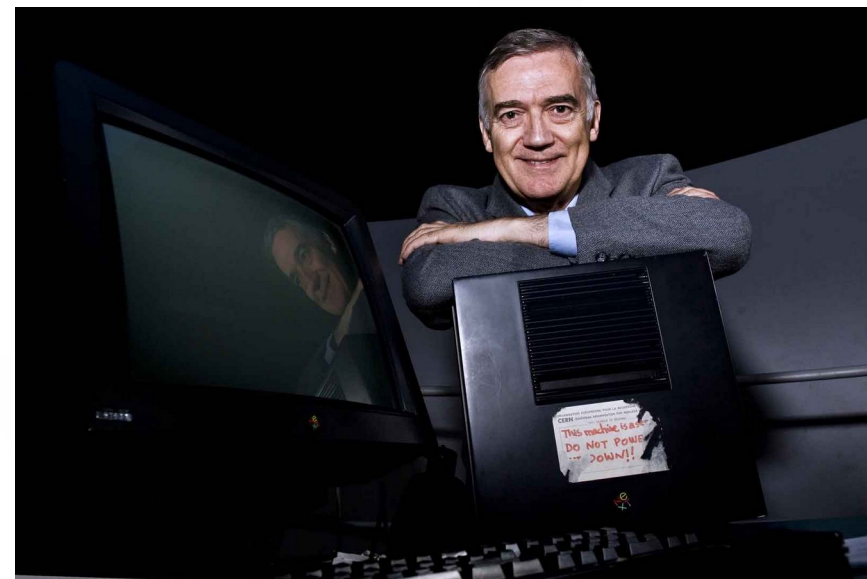


Histoire : naissance du WWW

- **Inventeur** : Tim(othy John) Berners-Lee, au CERN
- **13/03/1989** : Proposition système hypertexte organisé en toile
- **1990** : Robert Cailliau rejoint le projet Développent les **3 principales technologies** :
 - Adressage Web : **URL**
 - Protocole de transfert : **HTTP**
 - Langage des pages : **HTML** (**H**yper**T**ext **M**akup **L**angage)
- **1991** : Premier navigateur en mode texte
- **06/08/1991** : Première page
- **1993** : MOSAIC, premier navigateur graphique
- **30/04/1993** : WWW dans le domaine public



Tim Berners-Lee, CERN ©

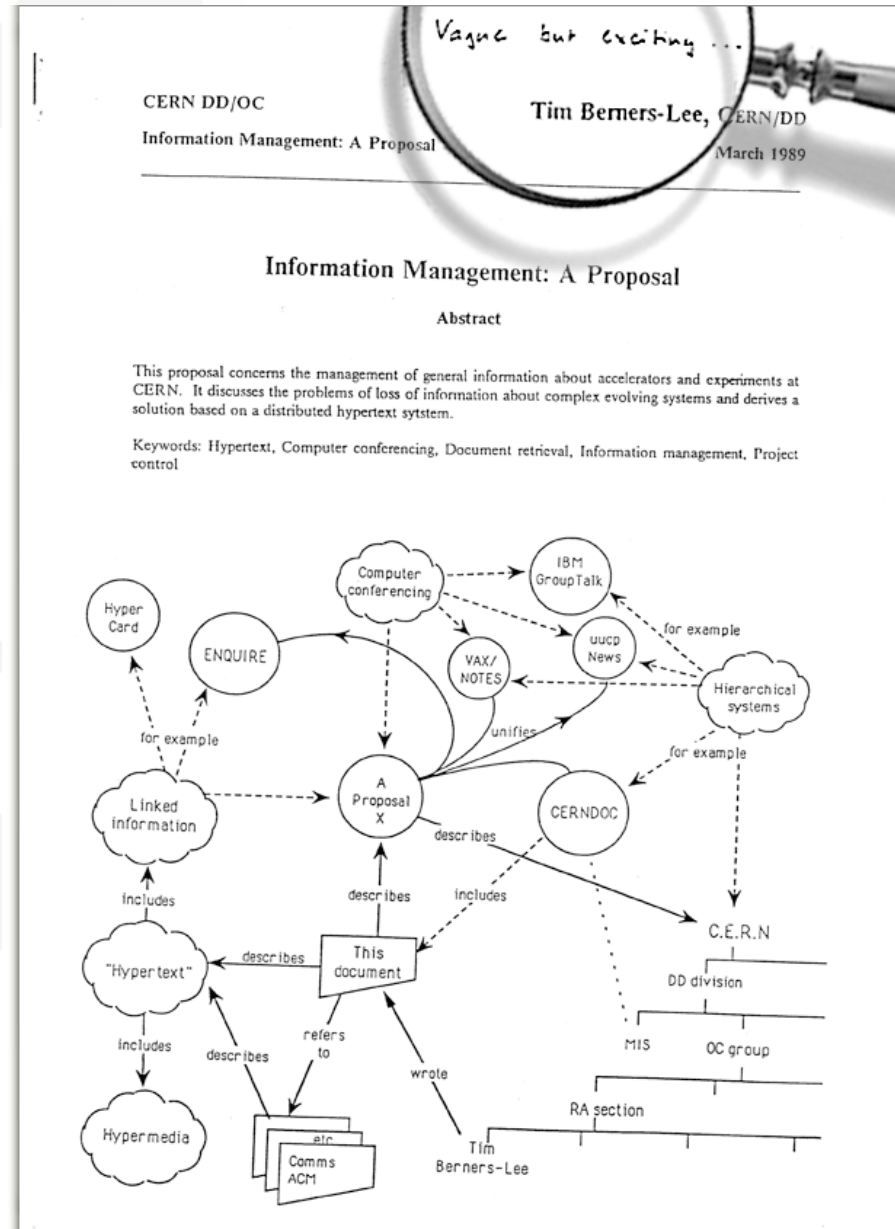


Robert Cailliau, CERN ©





1989 : Proposition de Tim Berners-Lee



Le premier serveur Web



Navigateurs



```
CERN CERN Welcome

The European Laboratory for Particle Physics, located near Geneva[1] in
Switzerland[2] and France[3]. Also the birthplace of the World-Wide
Web[4].

This is the CERN laboratory main server. The support team provides a set of
Services[5] to the physics experiments and the lab. For questions and
suggestions, see WWW Support Contacts[6] at CERN

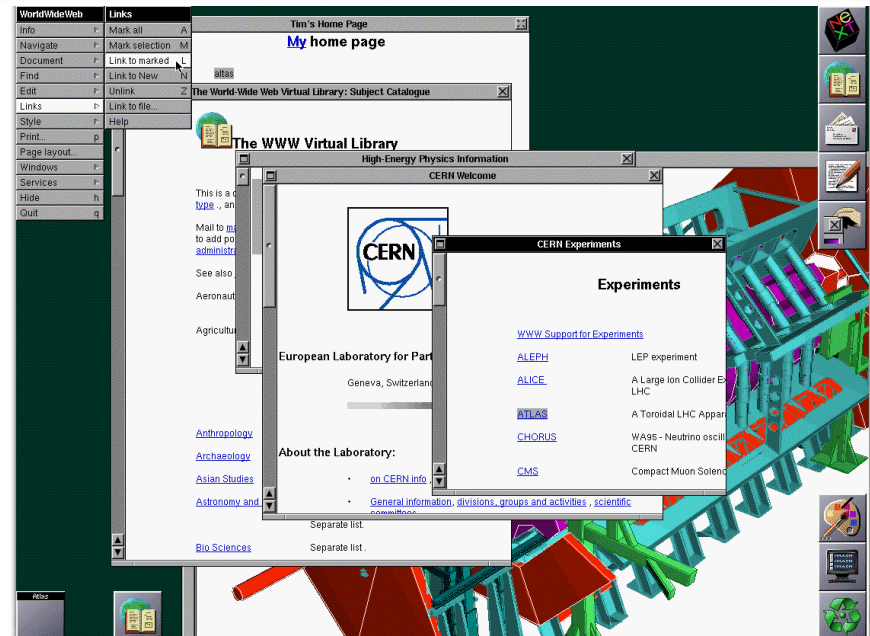
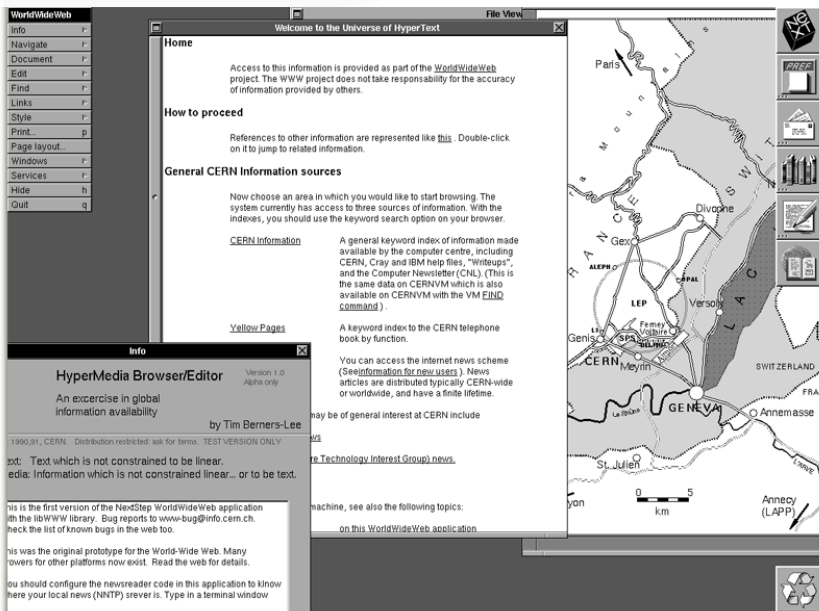
About the Laboratory[7] - Hot News[8] - Activities[9] - About Physics[10] -
Other Subjects[11] - Search[12]

About the Laboratory

Help[13] and General information[14], divisions, groups and
activities[15] (structure), Scientific committees[16]

Directories[17] (phone & email, services & people), Scientific
Information Service[18] (library, archives or Alice), Preprint[19] Server

1-45, Back, Up, <RETURN> for more, Quit, or Help: █
```





Histoire : quelques dates

● **1994** : Consortium **W3C**, Yahoo!, navigateur Netscape.

● **1995** : Tentative de Microsoft : réseau MSN.



● **1995** : Serveur HTTP Apache, Internet Explorer.

● **1995** : Moteur de recherche AltaVista, Hotbot.



● **1995** : Java, Javascript, PHP.



● **1996** : Feuilles de style : CSS.

● **1998** : Mozilla, Création de Google.



● **2001** : Wikipédia.

● **2003** : Création de **Web 2.0** : contenu généré par l'utilisateur.

● **2004** : Facebook, Firefox.



● **2006** : Twitter.

● **2007** : HTML 5, **2008** : Chrome.

● **2010** : Instagram, **2011** : Snapchat.



● **2015** : Microsoft Edge.





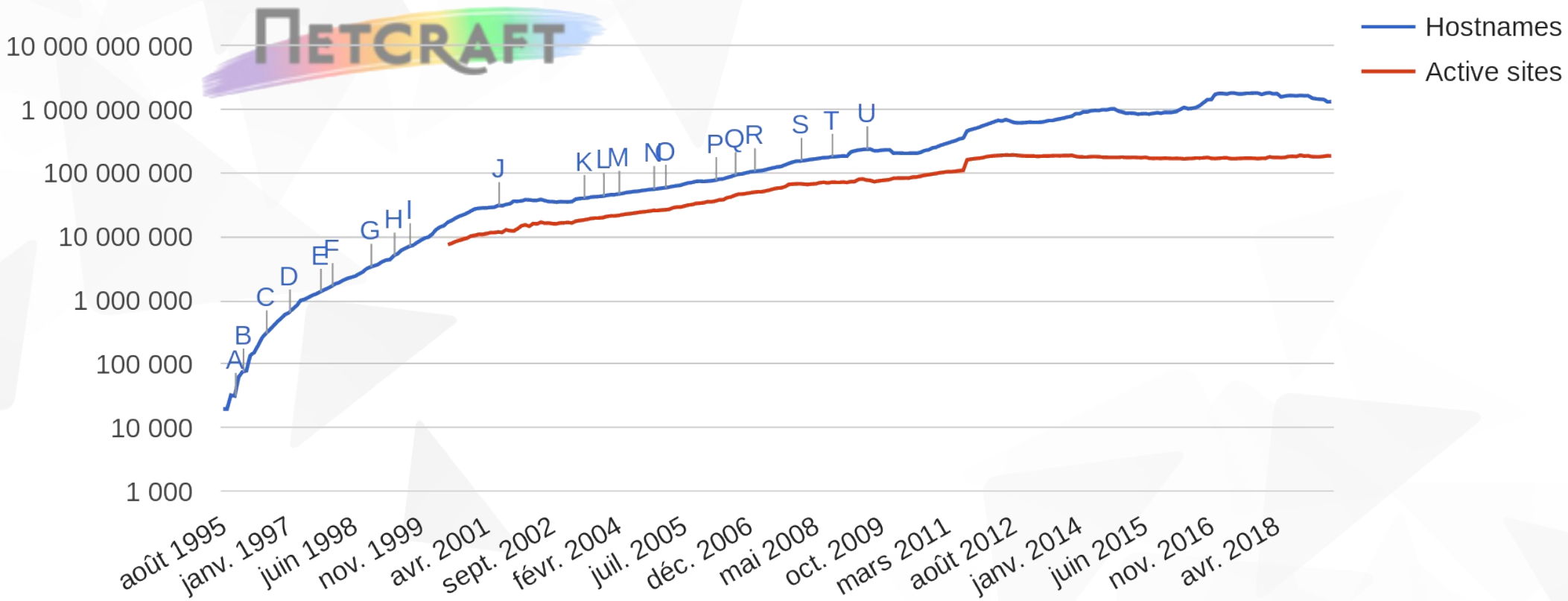
Évolutions

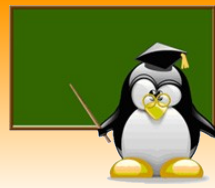
- Nombre de sites
- Classements
- Deep – Dark Web

Evolution du nombre de sites (log)



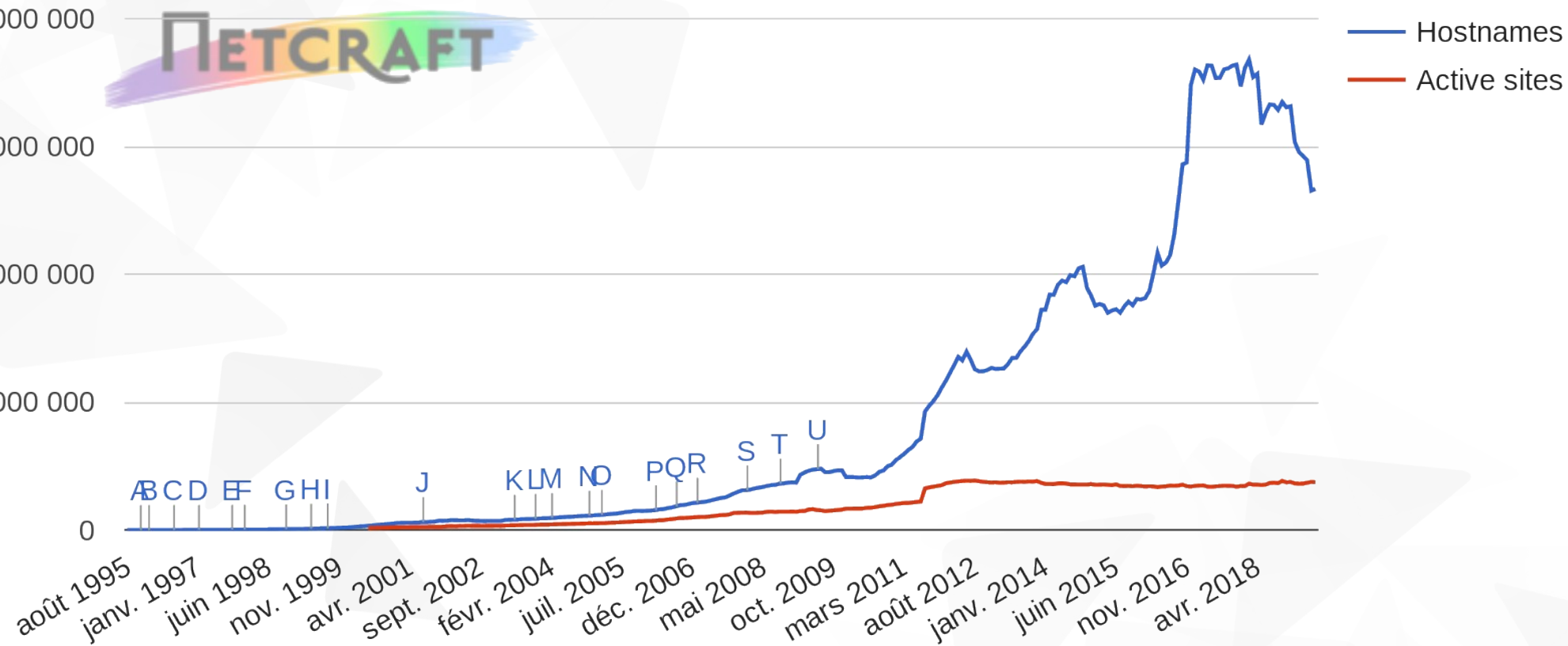
Total number of websites (logarithmic scale)



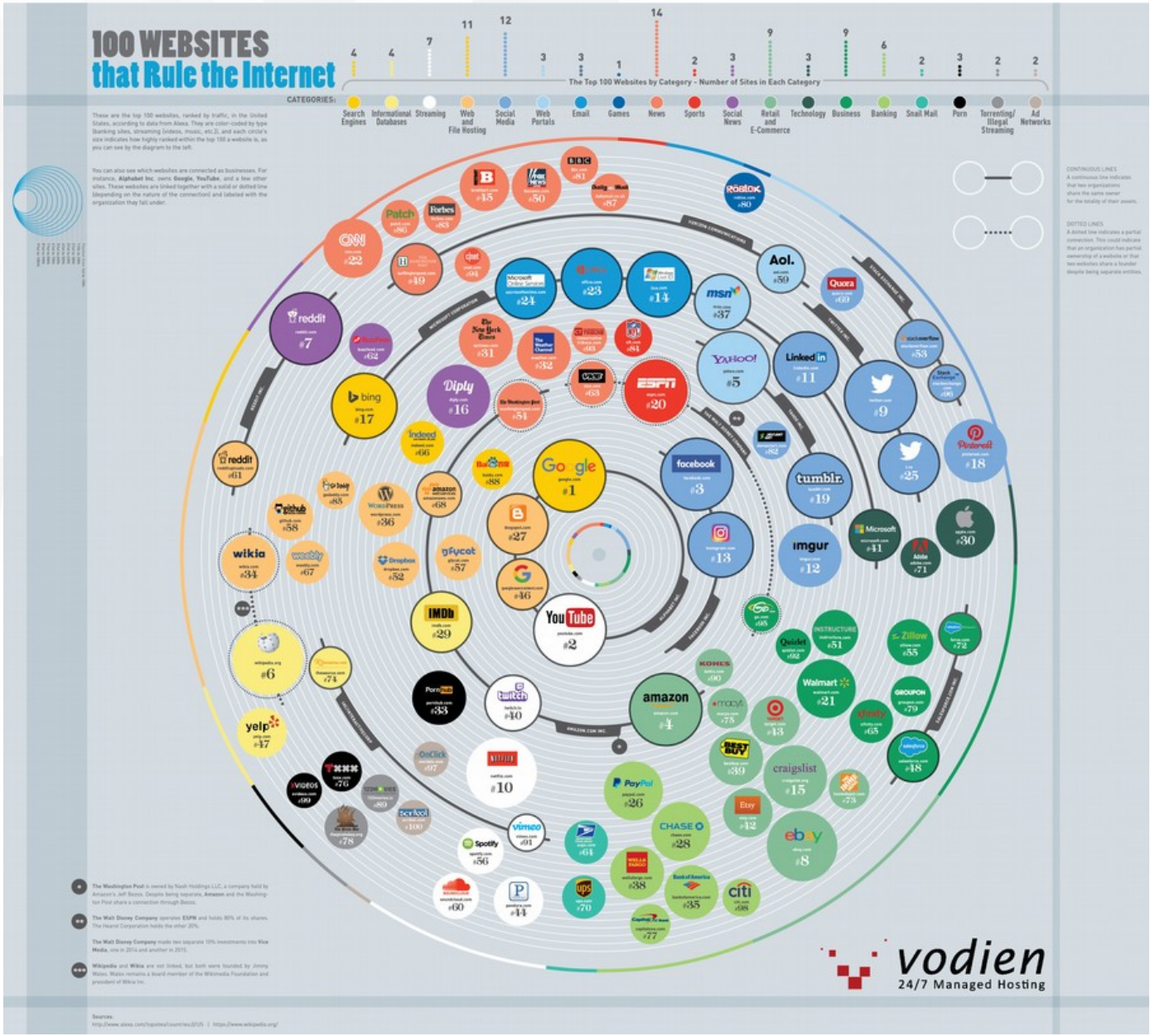
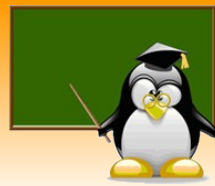


Evolution du nombre de sites (linéaire)

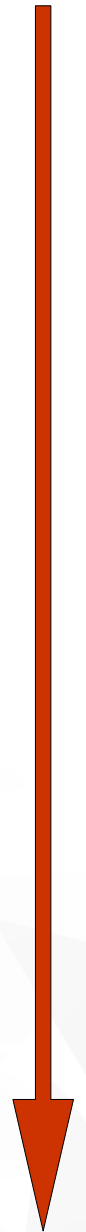
Total number of websites (linear scale)



Classement des sites



- **Classement :**
- Google
- YouTube
- Facebook
- Amazon
- Yahoo!
- Wikipédia
- Reddit
- Ebay
- Twitter
- Netflix
- ...



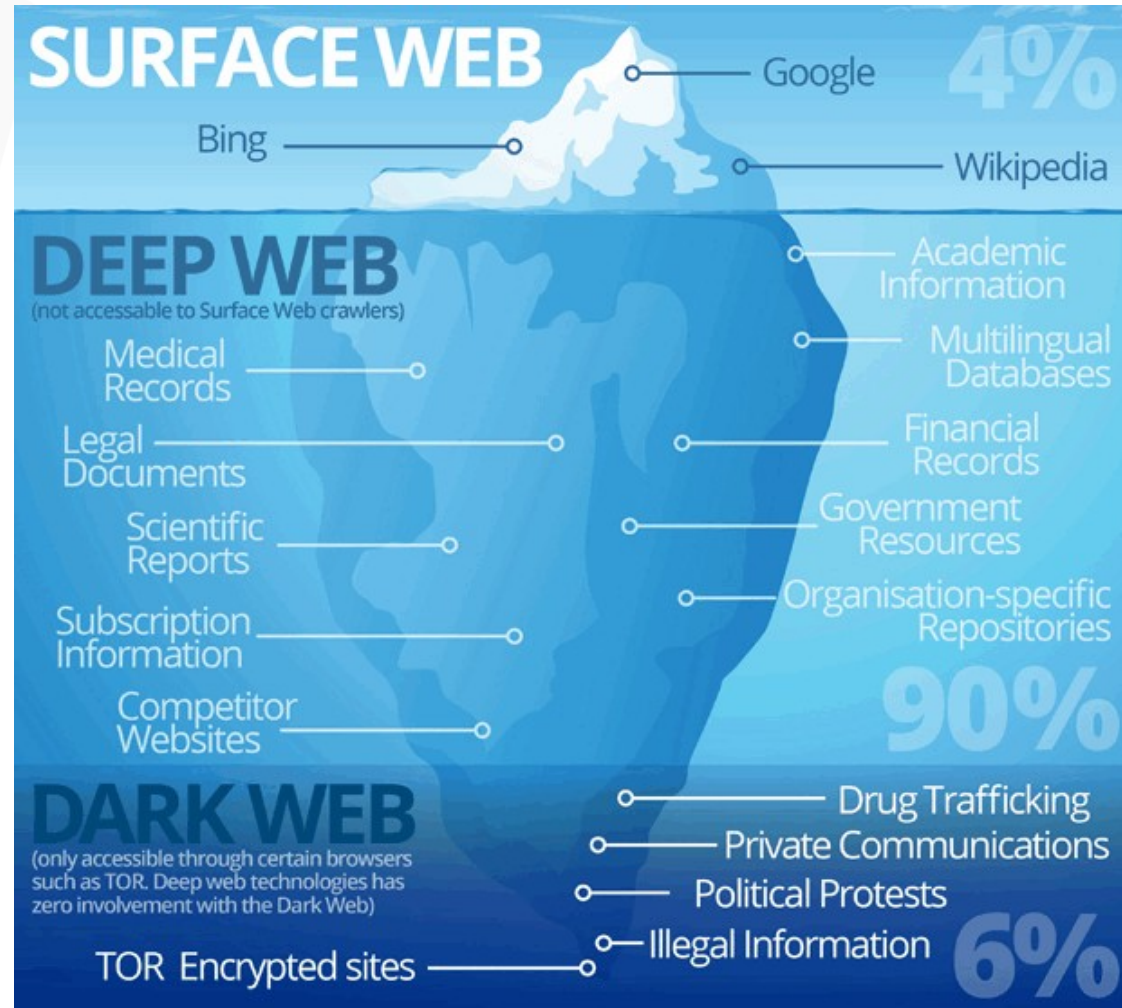


Deep Web – Dark Web

- Aussi appelé Web **Caché**, **Sombre** : face cachée d'Internet.
- Représente **96 %** du Web ...

Plusieurs **niveaux** :

- **Commun** : ce qui est visible de tous : YouTube, Facebook, Twitter, etc.
- **Surface** : sites encore visibles : reddit, hébergeurs web, bases de données.
- **Bergie** : dernier niveau encore accessible librement.
- **Deep** : commence à être obscure, pages non indexées.
- **Marianas** : réservé aux experts.





Une illustration ... des profondeurs





Fonctionnement

- Consultation page Web



Consultation page Web

- Navigateur sur un adresse (URL), par exemple :
<http://www.wikipedia.org>
- Ce qui se passe :
 - 1) On saisit une **adresse** Web
 - 2) Recherche du **serveur** : nom par rapport à l'adresse IP (DNS)
 - 3) Ce serveur a une **adresse IP** (cf. plus tard) unique et publique
 - 4) Serveur **trouvé** : demande le contenu d'une page (accueil)
 - 5) Le serveur **répond** à notre requête
 - 6) Notre navigateur **affiche** le résultat de la requête





Saisie adresse

Applications Google Chrome jeu. 18 juil., 10:23

Nouvel onglet x +

← → G

☆

Gmail Images

Google

Effectuez une recherche sur Google ou saisissez une URL

M Connexion CEREEP Chrome Web S... Ajouter un rac...

Personnaliser

Nouvel onglet - Google Chrome 1 / 4

Zone de saisie de l'adresse de la page à consulter : **URL**





Réponse du serveur

Applications Google Chrome jeu. 18 juil., 10:24

W Wikipedia +

← https://www.wikipedia.org

Adresse de la page : URL

Titre et icône de la page

Contenu de la page

WIKIPEDIA
L'encyclopédie libre

Français 2 123 000+ articles

English 5 892 000+ articles

日本語 1 159 000+ 記事

Español 1 532 000+ artículos

Deutsch 2 323 000+ Artikel

Русский 1 556 000+ статей

Italiano 1 541 000+ voci

中文 1 065 000+ 條目

Português 1 010 000+ artigos

Polski 1 346 000+ haset

FR

Lire Wikipédia dans votre langue

Wikipédia est hébergé par la Wikimedia Foundation, un organisme sans but lucratif qui héberge également beaucoup d'autres projets.

Les applications Wikipédia sont désormais disponibles :
Téléchargement pour iOS sur l'App Store
Téléchargement pour Android sur Google Play
Voir la liste complète des applications
Wikipédia disponibles

Commons Photos et autres librement utilisables

Wikilivres Textes pédagogiques libres

Wikiversité Matériel de cours libre d'utilisation

Wikisource Bibliothèque libre

Wikivoyage Guide de voyage gratuit

Wikinews La source d'informations que vous pouvez écrire

Wikiquote Collection libre des citations

Wikispecies Répertoire libre des espèces

Wiktionnaire Le dictionnaire libre

Wikidata Base de connaissance libre

MediaWiki Application wiki libre et ouverte

Meta-Wiki Coordination et documentation de la communauté

Wikipedia - Google Chrome 1 / 4





Détails de la réponse à la requête

Applications Google Chrome | jeu. 18 juil., 10:25

Wikipedia | https://www.wikipedia.org

WIKIPÉDIA
L'encyclopédie libre

Français 2 123 000+ articles | English 5 892 000+ articles
日本語 1 159 000+ 記事 | Español 1 532 000+ artículos
Português 1 010 000+ artigos | Polski 1 346 000+ hasel

Code source de la page Web reçue suite à notre requête

```
<html lang="fr" class="js-enabled" dir="ltr">
<head>
  <meta charset="utf-8">
  <title>Wikipedia</title>
  <meta name="description" content="Wikipedia is a free online encyclopedia, created and edited by volunteers around the world and hosted by the Wikimedia Foundation.">
  <script></script>
  <meta name="viewport" content="initial-scale=1,user-scalable=yes">
  <link rel="apple-touch-icon" href="/static/apple-touch/wikipedia.png">
  <link rel="shortcut icon" href="/static/favicon/wikipedia.ico">
  <link rel="license" href="//creativecommons.org/licenses/by-sa/3.0/">
  <style></style>
  <link rel="preconnect" href="//upload.wikimedia.org">
  <link rel="preconnect" href="//fr.wikipedia.org">
</head>
<body id="www-wikipedia-org" class="jsl10n-visible" == #0
  <h1 class="central-textlogo" style="font-size: 1em; title="Wikipedia">
  <!-- container div for the central logo and the links to the most viewed language editions -->
  <div class="central-featured" data-el-section="primary links"></div>
  <!-- Search form -->
  <div class="search-container"></div>
  <div class="lang-list-button-wrapper"></div>
  <div class="lang-list-border"></div>
  <div class="lang-list-container"></div>
  <hr>
  <div class="footer" data-el-section="other projects"></div>
</body>
```

Wikipédia est hébergé par la Wikimedia Foundation, un organisme sans but lucratif qui héberge également beaucoup d'autres projets.

Commons Photos et autres librement utilisables

Wikivoyage Guide de voyage gratuit

Wiktionnaire Le dictionnaire libre

Wikilivres Textes pédagogiques libres

Wikinews La source d'informations que vous pouvez écrire

Wikidata Base de connaissance libre

Wikiversité Matériel de cours libre d'utilisation

Wikiquote Collection libre des citations

MediaWiki Application wiki libre et ouverte

Wikisource Bibliothèque libre

Wikispecies Répertoire libre des espèces

Meta-Wiki Coordination et documentation de la communauté

Wikipedia - Google Chrome | 1 / 4





Plus de détails : Réseau

- Nom de domaine
- Adresse IP



Nom de domaine : par ex. « upupides.fr »

- **Identifie** une machine (ou +) sur le **réseau**
- **Racine** du nom du serveur
- **Unique, déposé** auprès d'un « registrar »
- Centralisation : **NIC** « **N**etwork **I**nformation **C**enter »
 - **AFNIC** : pour la France
 - CIRA, DENIC, VeriSign Global Registry, EURid
- Nom de domaine composé de :
 - **Label** : '*upupides*'
 - **TLD** → « **T**op **L**evel **D**omain » : '*fr*'
 - TLD **réservés** : 'gov', 'mil'
 - TLD **libres** : 'biz', 'com', 'io', 'net', 'org', 'pro', 'fr', etc.
 - **+ ou -** cher ...



afnic





Nom de domaine

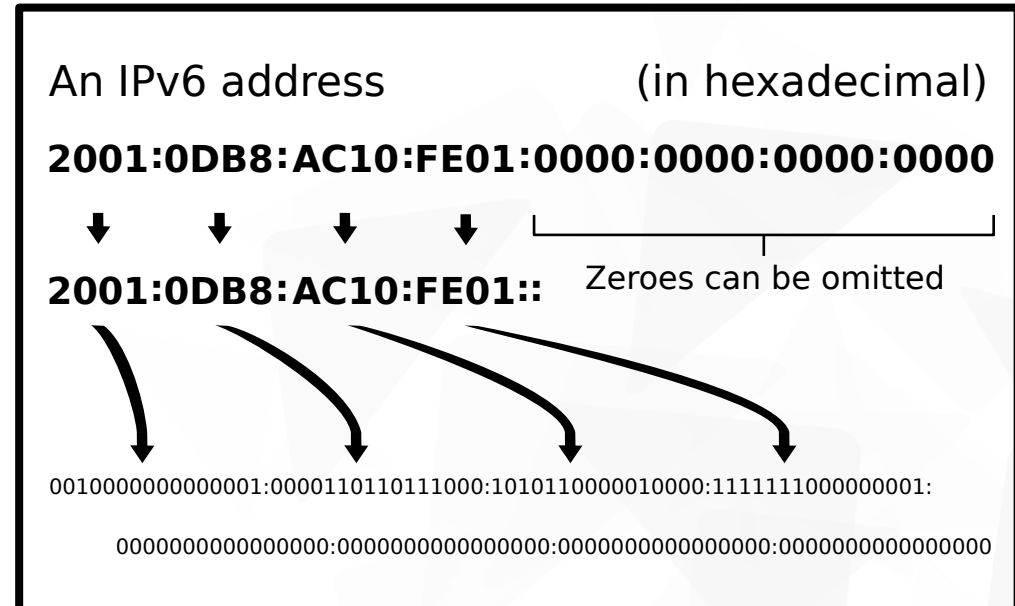
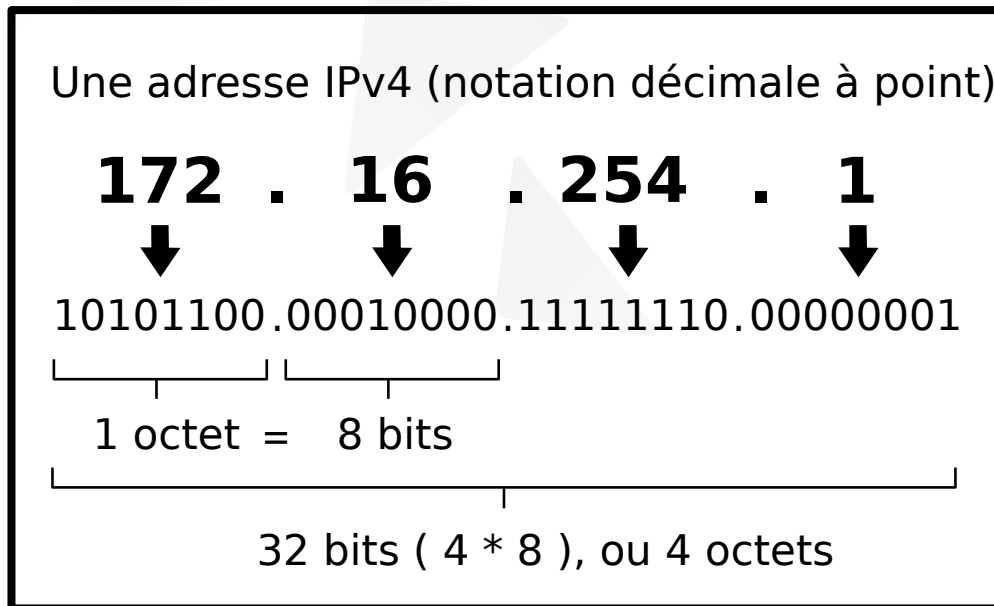
- Adresses : <http://www.upupides.fr> ou : <http://111.222.333.444>
- '**http**', '**https**', '**ftp**', etc. : protocole de communication
- ':' : caractère séparateur
- '**//**' : marqueur de début de chemin
- '**www**' : l'hôte, identifiant machine
- '**upupides**' : label (max = **63 caractères** - longueur TLD)
- '**fr**' : TLD (Top Level Domain)
- Label et hôte : de '**a/A**' à '**z/Z**', '**0**' à '**9**', et '**-**' (sauf début et fin)
- **3 identités** par domaine : Administrative, Billing, Technical
- **Informations** visibles en ligne : <http://www.whois.net>
- **DNS** = **D**omain **N**ame **S**erver.
- DNS : association **adresse IP** ↔ **nom de domaine**





Adressage IP

- Serveur **DHCP** = **D**ynamic **H**ost **C**onfiguration **P**rotocol : configuration **automatique** des interfaces réseau
- **2 types** d'adresse IP :
 - **Statique** : toujours la même adresse IP
 - **Dynamique** : adresse affectée selon disponibilité
- Par exemple : 'www.google.fr' → 108.177.126.94
- Configuration **aisée** si adresse **IP fixe**
- Service du type '**DynDNS**' pour adresse flottante : <http://www.dyndns.org>





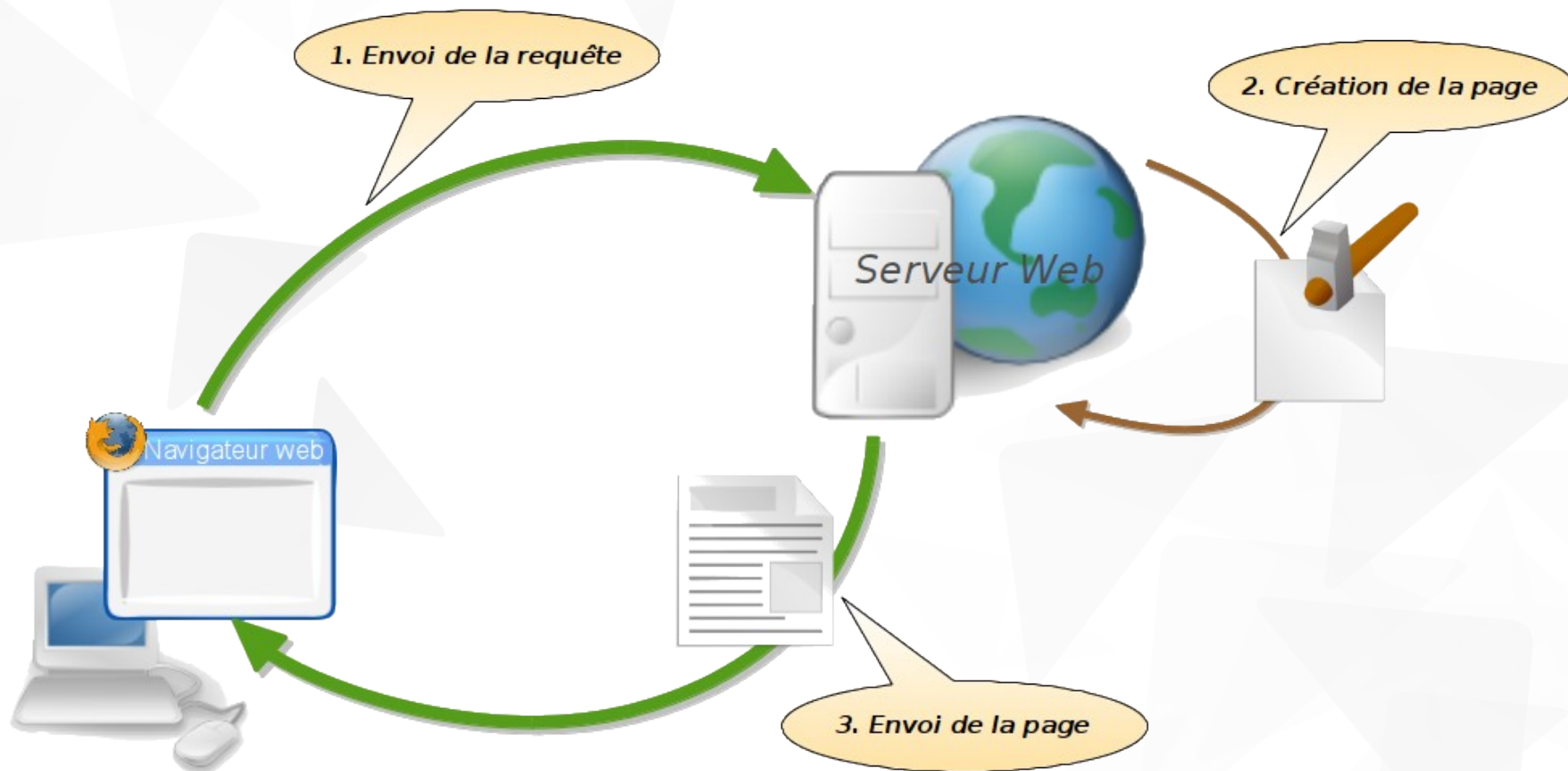
Plus de détails : Serveurs Web

- Fonctionnement
- Statique vs Dynamique
- Solutions disponibles



Serveur Web : fonctionnement

- Application qui **interprète** et **répond** à des **requêtes** HTTP
- Exécutée sur une machine accessible par **réseau**
- Permet à des **clients** (navigateur) d'accéder à du **contenu**
- Pages Web : **résultats** des requêtes au format HTML





Serveur Web : Statique vs Dynamique

1 Requête du client

3 Envoi HTTP du fichier



2 Lecture du fichier

1 Requête du client

3 Envoi HTTP du contenu



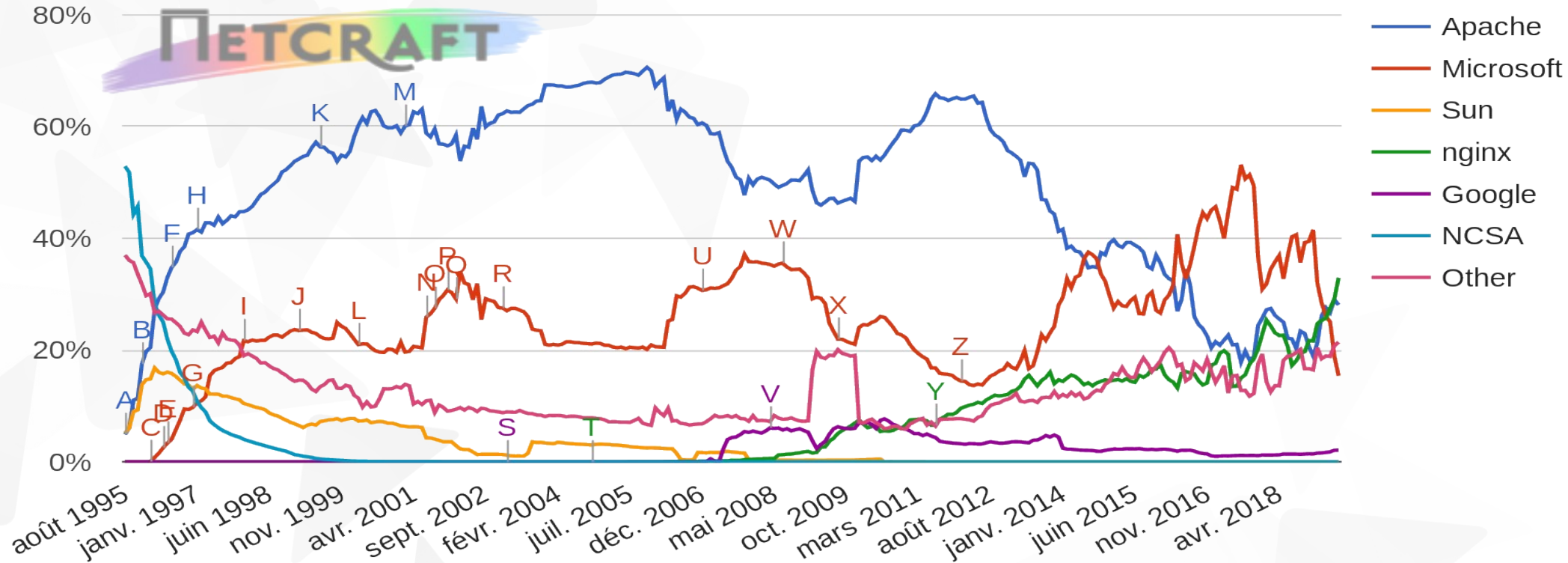
2 Exécution de l'application
et lecture de la sortie standard





Serveur Web : Solutions disponibles

Web server developers: Market share of all sites



Produit	Dév.	Sites hébergés	%
NGINX	Igor Sysoev	439 626 713	32,97 %
Apache	Apache	374 360 949	28,08 %
IIS	Microsoft	205 235 291	15,39 %
GWS	Google	28 181 744	2,11 %





Hébergements

- Hébergement gratuit
- Hébergement mutualisé
- Hébergement dédié
- Résumé



Hébergement gratuit

- « Quand c'est gratuit, le produit, c'est vous »
- Proposé par de nombreux **FAI** (**F**ournisseur d'**A**ccès **I**nternet)
- Espace de stockage **limité** : 10 Mo..1 Go
- Temps d'accès assez **lent**
- Pas de possibilité de régler le nom de domaine
- Disponibilité et sauvegarde **non garanties**
- Souvent : **bannières** et **publicités** insérées
- Transfert par protocole **FTP** = **F**ile **T**ransfert **P**rotocol
- **Une** seule base de données, quand il y en a ...
- Fonctionnalités (PHP, par ex.) assez **réduites**
- Convient pour un **usage personnel**, simple





Hébergement mutualisé

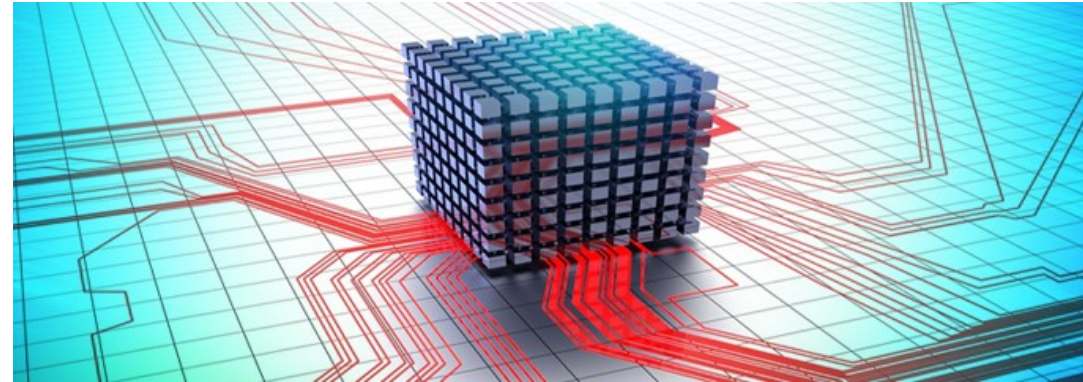
- **Plusieurs sites** sont hébergés sur un **seul et même** serveur
- Espace de stockage **plus important**
- **Meilleure** qualité de service coté maintenance système
- Offre de **sauvegarde** et **restitution** des données
- Pas d'accès direct Administrateur
- Configuration par **page Web**
- Prix **proportionnel** :
 - Au nombre de bases de données
 - A l'espace occupé
 - Aux nombres de visiteurs potentiels
- **Outils** d'analyse et de statistiques intégrés
- Installation **centralisée**, conflits de versions possibles
- Domaines et sous-domaines **disponibles**





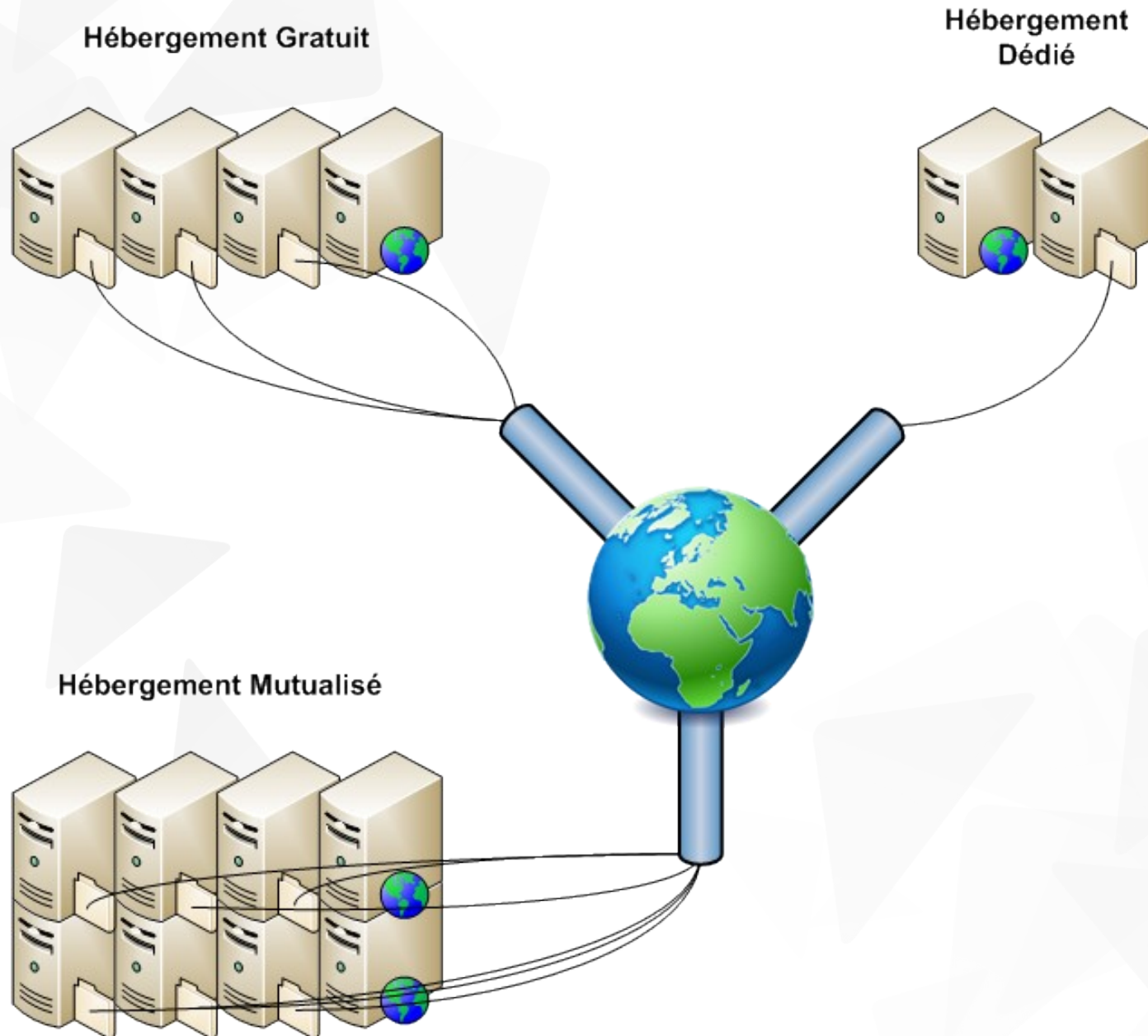
Hébergement dédié

- Le plus **confortable**
- Le plus **sûr**
- Le plus **cher** ...
- Contrôle **total** de la machine :
 - Vous êtes maître à bord
 - Administrateur du serveur
 - Installation possible de machines virtuelles, de bureau à distance, etc.
- **Connaissances** poussées en administration système et réseau
- **Indépendance** sur les applications
- Mise à jour **régulière** des applicatifs nécessaires





Offres d'hébergement





Résumé offres d'hébergement

- **Gratuit** :
 - Rapide à utiliser, fonctionnalités basiques, fiabilité réduite
 - Pour les sites personnels, ou pour tester
- **Mutualisé** :
 - Bon rapport qualité / prix. Bonnes capacités de stockage
 - Outils intégrés et installation rapide
 - Conflits de versions et problèmes de charge possibles
 - Pour les sites sans gros trafic, ni surcharge
- **Dédié** :
 - Le plus coûteux
 - Compétences techniques non négligeables
 - Pour les sites professionnels





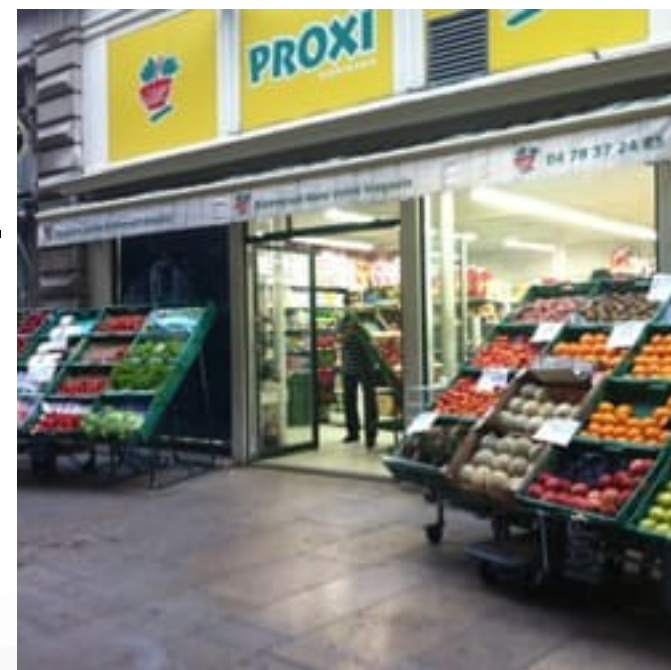
Serveurs particuliers

- Proxy
- Base de données (BDD)



PROXY : Fonctionnement

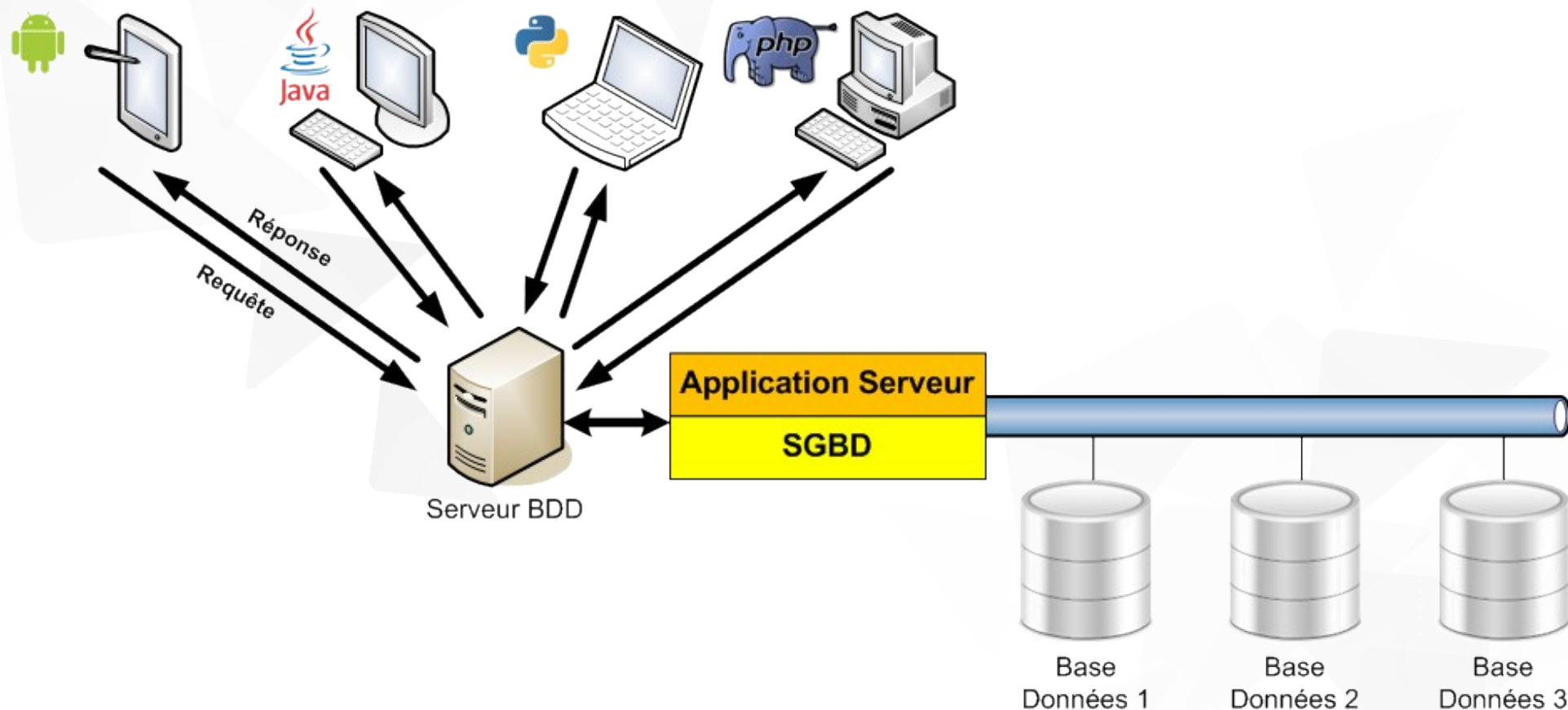
- Serveur **intermédiaire** entre une machine et Internet
- Permettent de **stocker** en **cache** certaines pages Web
- Temps d'affichage et d'accès améliorés
- **Renforce** la sécurité :
 - **Filtrage** possible : adresses, popup, etc.
 - **Authentification** possible
 - **Journalisation** des accès
 - **Contrôle** parental
- Possible de devenir « **anonyme** » :
 - Seul le proxy est visible de l'extérieur





Bases de données : Fonctionnement

- Service **structuré** de stockage de données avec minimum de redondance
- Accessibles par des **applications** et **utilisateurs** différents
- **SGBD** = **S**ystème de **G**estion de **B**ases de **D**onnées





Bases de données : principes

- Une **base de données** inclut plusieurs **tables**
- Comme un **classeur** avec plusieurs **feuilles**
 - **Classeur** = Base de données
 - **Feuille** = Table
- Ses **rôles** :
 - **Accéder** aux données d'une façon simple
 - Autoriser l'accès à **plusieurs** utilisateurs / applications
 - **Manipuler** les données dans la base
 - Mise à jour / consultation de données avec **droits d'accès**
 - **Accès** multi-utilisateurs **simultané** possible
 - Plusieurs **types** : MySQL, SQLite, PostgreSQL, mSQL, etc.





Structure et sécurité

- Structure et arborescence
- Protocole HTTP
- Norme CGI
- DMZ



Structure, Arborescence d'un serveur Web

- **Structurer** site Web = **ranger** les choses à leur place
- Quelques **exemples** possibles :
 - **'private'** : privé, avec droits d'accès, protéger données sensibles
 - **'images'** : public pour incruster des images
 - **'upload'** : téléchargement de fichiers
- Serveur Web fait le **lien** entre une **URL** et un **point d'entrée** de l'arborescence située sur le disque :
Ex : <http://www.upupides.fr/images> → C:\www\images
- **Attention** au répertoire d'entrée réglé → possibilité de donner l'accès à des répertoires système
- Plus facile de **sécuriser** un répertoire qu'un seul fichier





Protocole HTTP

- **Communication** entre navigateur et serveur
- Actuellement : **HTTP/1.1** : décrite dans la doc « **RFC 2616** »
- **IETF** = The **I**nternet **E**ngineering **T**ask **F**orce
- **Codes** d'exécution :

http://

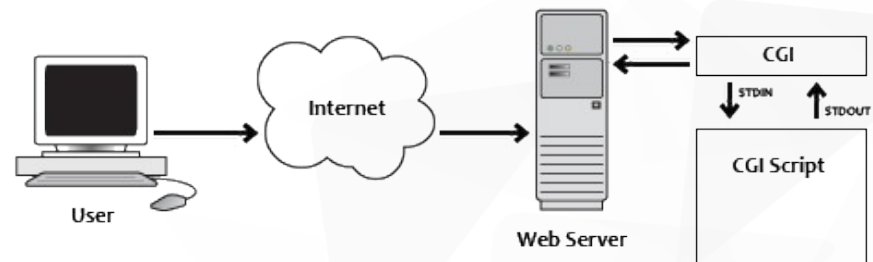
CODE	Statut	Description
200	OK	Requête traitée, sans erreur
203	Partial Information	Requête traitée, informations retournées incomplètes
204	No Response	Requête traitée, mais pas de réponse au client
400	Bad Request	Requête incorrecte
403	Forbidden	Requête traitée, problème de permissions d'accès aux données
404	Not Found	Page non trouvée
500	Internal Error	Erreur interne au serveur
501	Not Implemented	Manque la fonctionnalité pour traiter la requête sur le serveur





Norme CGI

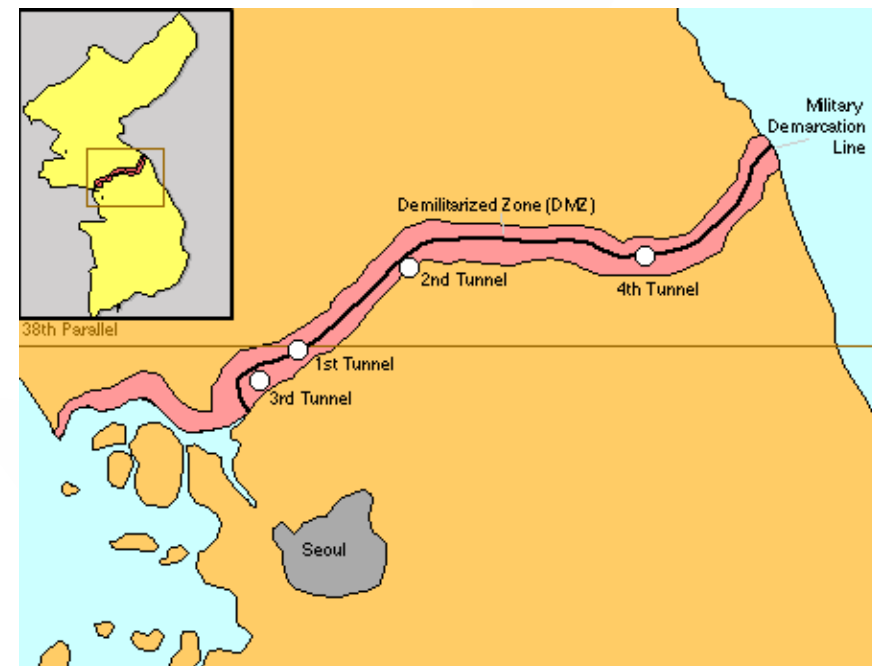
- **CGI** = **C**ommon **G**ateway **I**nterface
- Version 1.1 (octobre 1995)
- Au début beaucoup de pages statiques, avec des liens
- Besoin rapide de « **dynamiser** » le contenu
- Affichage du **résultat** (pas le contenu) de l'exécution d'un programme
- Obligation d'envoyer les **entêtes** HTTP dans le programme
- **Adresse** Web du type : `http://www.upupides.fr/cgi/test`
- Plusieurs déclinaisons :
 - **FCGI** = **F**ast **C**GI
 - **SCGI** = **S**imple **C**GI
- Possibilité de faire tourner plusieurs **types** de langages :
 - Python, C, C++, Java, etc.





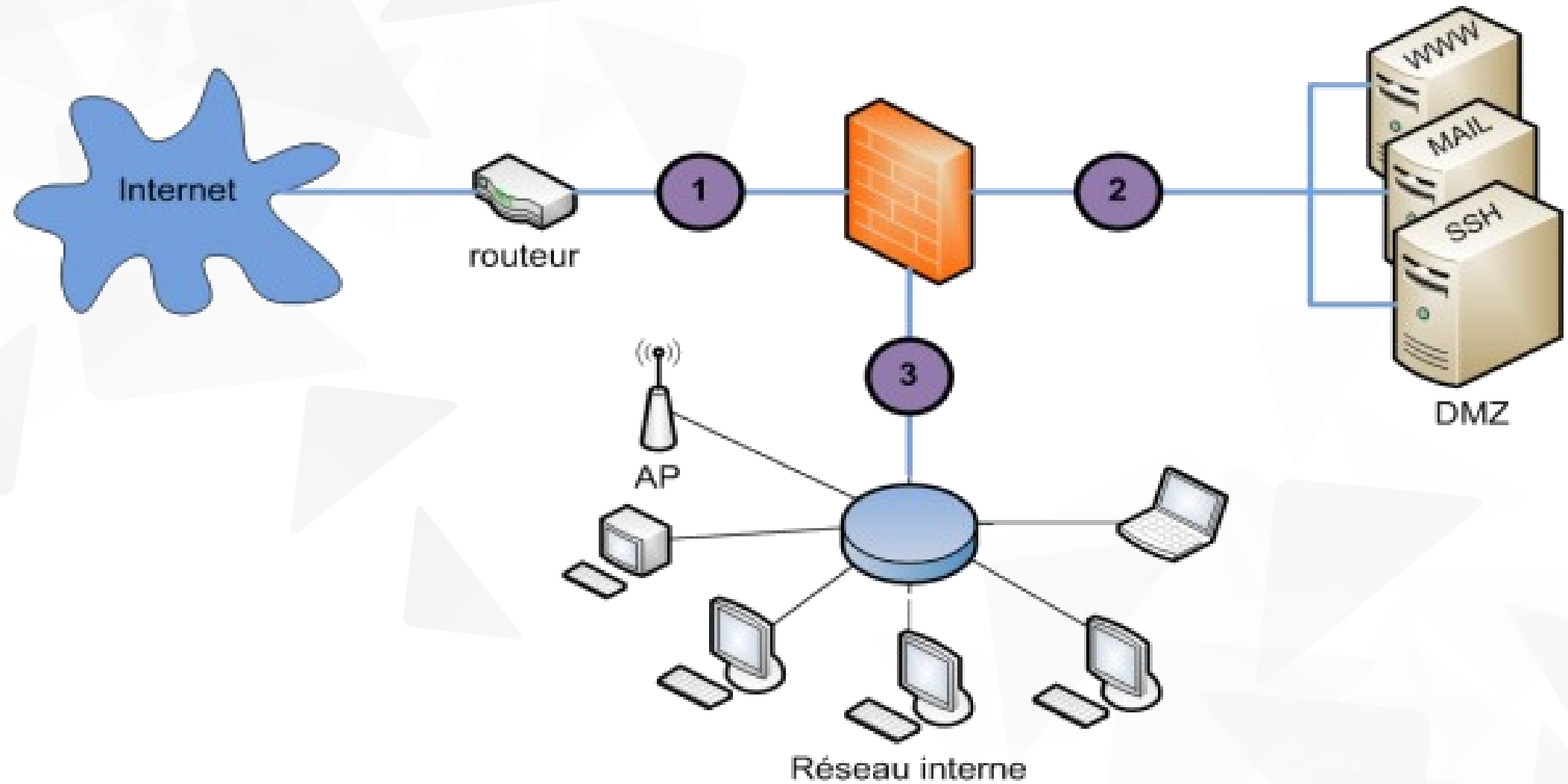
DMZ : Fonctionnement

- **De**Militarized **Z**one : à l'origine zone « franche » entre la Corée du Nord et la Corée du Sud
- Sous-réseau **séparé** du réseau local
- **Pare-feu** (« Firewall ») qui isole les réseaux
- Contient **machines** pouvant être accédées depuis **Internet**
- Pare-feu **bloque** les accès au réseau local
- Pirates **cloisonnés** dans cette zone
- **Problème** si le pare-feu compromis
- Possible de **multiplier** les Firewall :
 - Pour plus de sécurité
 - Mais plus complexe à administrer



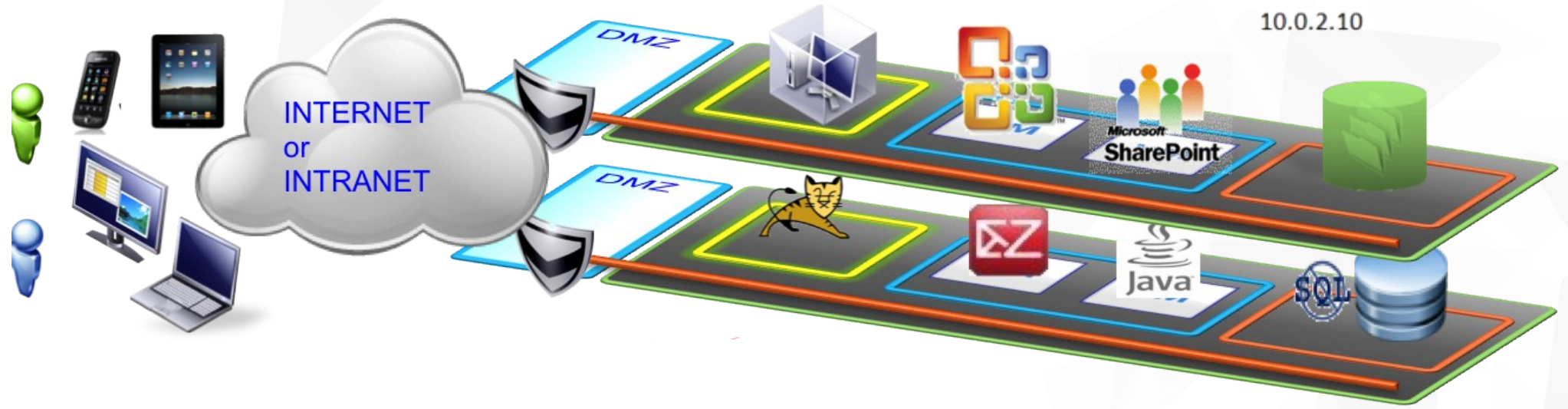
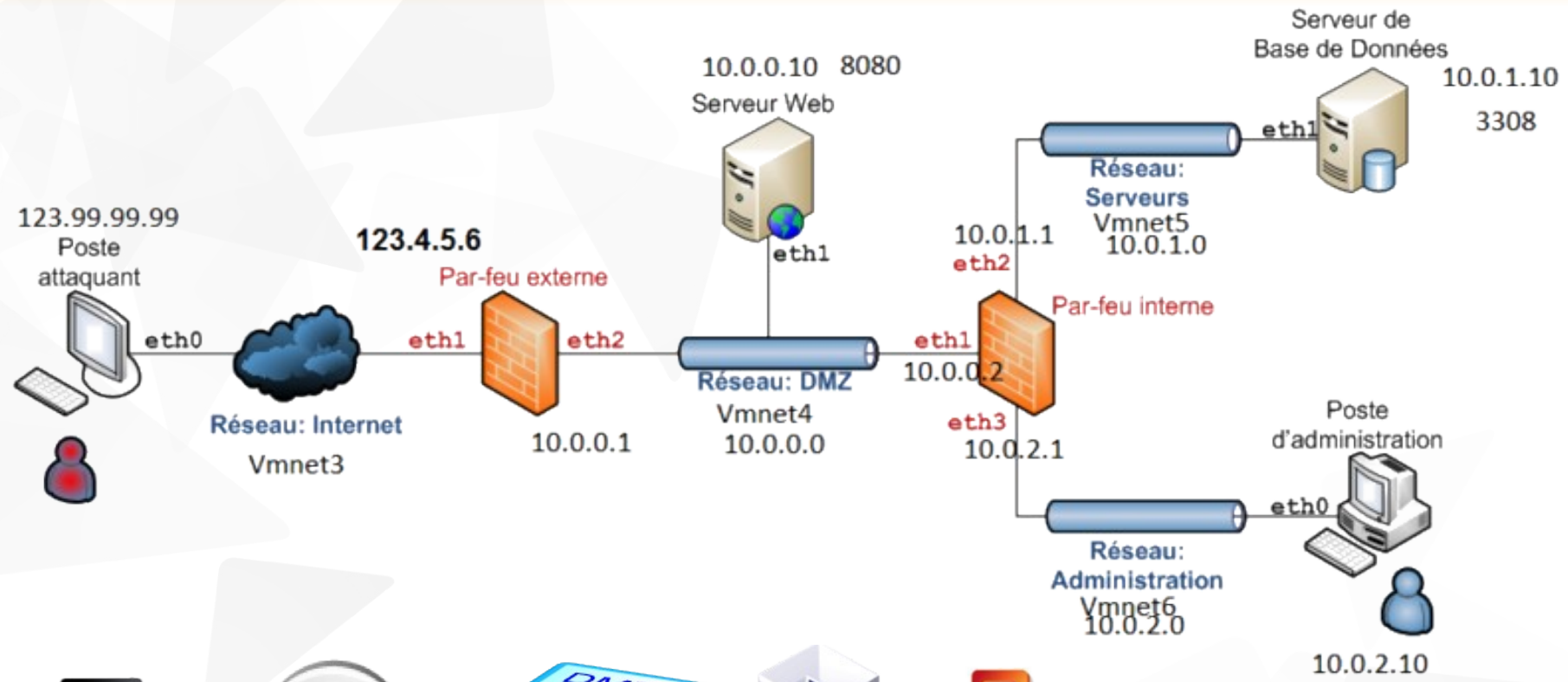


DMZ : Architecture de base





DMZ : Architecture « haute sécurité »





Conclusion

- Liens
- Conclusion



Liens

- Statistiques en **direct** :
<http://www.internetlivestats.com/>
- Statistiques **mensuelles** :
<https://www.netcraft.com/>
- Consortium **W3C** :
<https://www.w3.org/>
- Navigateur en **ligne de commande** :
<http://line-mode.cern.ch/www/help.html>
- **WHOIS** (AFNIC) :
<https://www.afnic.fr/fr/produits-et-services/services/whois/>





Conclusion

- Une technologie qui **relie** les **gens** mais aussi les **machines**
- **Grosse quantité** d'informations disponibles : **infobésité** ...
- **Internet** : une technologie qui évolue **très vite** :
 - Contenu
 - Fonctionnalités
 - Inter-connexions
- **Chacun** y participe de près ou de loin ...
- **Web 2.0** (2000..2009) : Web **social**, dynamisation, utilisateur participe au contenu
- **Web 3.0** (2010) : basé sur **expérience utilisateur**, structure masse d'informations selon localisation, les préférences, etc.
- **Web 4.0** : le futur (proche), des robots structurent ...
- **Principe de base** : sur Internet, tout est visible, et très certainement enregistré quelque part !



FIN !

Merci ...

- ... pour avoir tenu jusqu'ici,
- ... de ne pas hésiter à poser des questions !

