V1.7

Tests de montée en charge & Haute disponibilité

Appliqués à l'ENT de *Paris Descartes*

ESUP-Days 13

M. Lamaachi – D. Houdet Joly

8 Fév 2012

ESUP Portail

UNIVERSITÉ PARIS DESCARTES

Sommaire

- Contexte et enjeux à Paris Descartes
- Une architecture Apache/Tomcat en « load balancing »
- Tests de montée en charge pour sécuriser
- Vers une architecture à haute disponibilité
- Conclusion

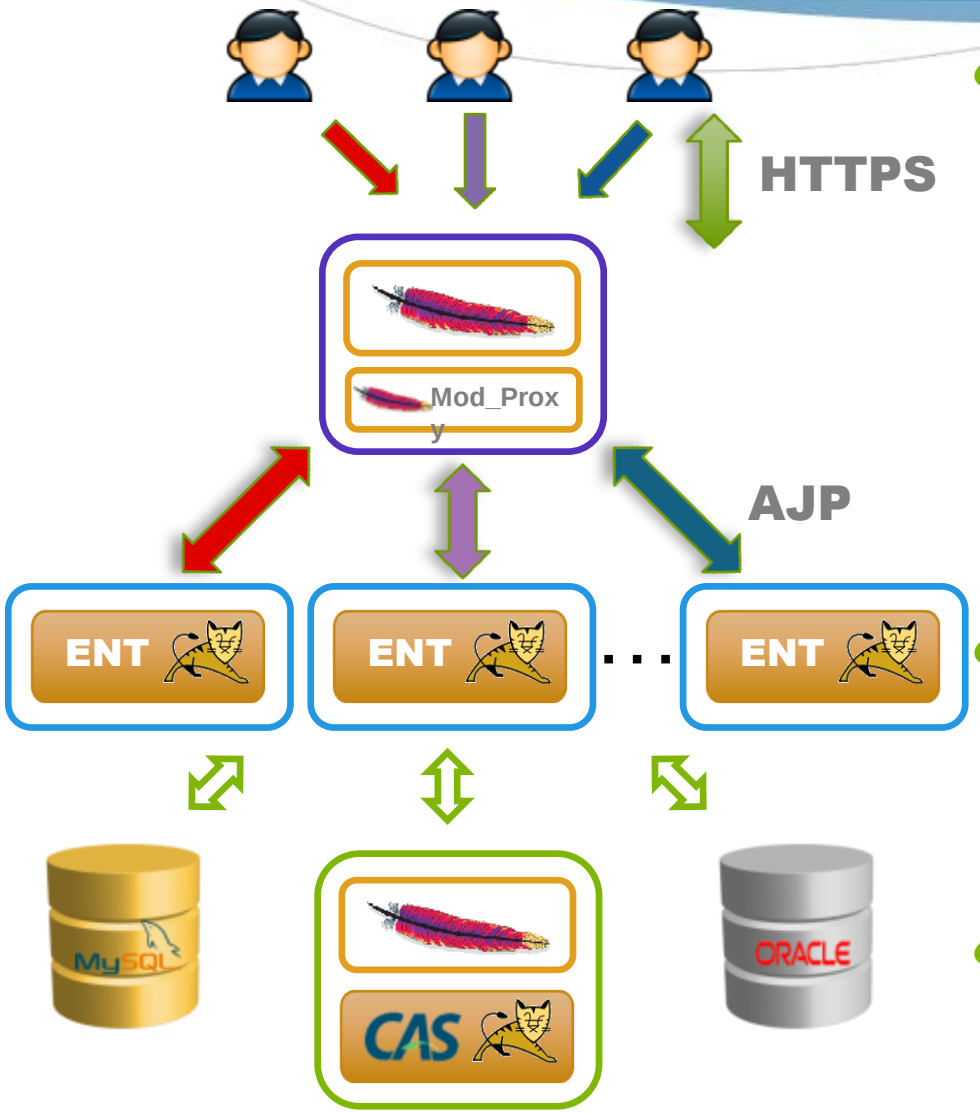
L'ENT à l'université Paris Descartes

- Université pluridisciplinaire
  - **répartie sur 9 Ufr (ou facultés) et 1 IUT**
- 37 000 étudiants
- 4000 personnels
  - **enseignants chercheurs, chercheurs, BIATSS**
- le point d'entrée unique vers les services numériques
  - **depuis fin 2008**
  - **au 15-12-2010 : 40 services dont 3 portlets**

Historique

- Installation de ESUP uPortal 2.6-dlm
- Montée en charge sans problème sur le début de l'année 2009
  - **en moy. 15000 consultations différentes par mois**
  - **35% des utilisateurs**
- Incidents bloquants (sept/oct 2009)
- Fin 2009, une nouvelle architecture est mise en place, préconisée par deux experts ESUP

Architecture opérationnelle en 2009



- Plus robuste
  - Architecture
    - **Frontal web** : serveur virtuel OpenVZ Debian - 256 Mo RAM - 4 Go disque
    - **Serveurs d'applications** : 10 serveurs virtuels VmWare Debian 32 bits - 2 Go RAM - 10 Go disque
  - Frontal
    - Serveur d'application protégé par le frontal apache mod_proxy
    - Load balancing + scripts de synchronisation
- Plus fiable
  - Sécurisation du socle et des applications
  - Monitoring des applications
    - Par Lambda Probe
- Plus évolutive
  - permet d'augmenter le nombre de serveurs d'application et de réaliser des migrations simplement

Les enjeux en 2010

- Organisation de PAES à partir de janvier 2010
  - **3500 étudiants qui accèdent à l'Ent en même temps**

- Des services plus consultés toute l'année
  - **en moy. 20 000 consultations différentes par mois**
  - **46% des utilisateurs potentiels**

Cette plateforme résistera-elle à un pic de charge?

- A quel moment
  - **En Janvier et en Juin, le jour de la publication des résultats PAES avec 3500 étudiants qui accèdent à l'ENT en même temps,**
  - **Les journées d'inscriptions**
- Pas moyen de le savoir, mais pour sécuriser…
  - **On augmente les backends pour passer de 10 à 20**
  - **On surveille tout les éléments de la plateforme**

Incident le jour de la publication des résultats 28 janvier 2011

- Surcharge du serveur Apache
  - **Le serveur web n'a pas résisté à la montée en charge, ce qui a rendu l'ENT inaccessible à plusieurs moments, en dépit des 20 backends.**
- Plantage du serveur CAS v2
  - **JavaOutOfMemory Exception**

Réunion de crise (évidement …)

- Objectif
  - **Réunir toutes les compétences pour discuter, analyser, et trouver des solutions**
  - **Explorer toutes les pistes susceptibles d'apporter des solutions, ainsi que les contraintes de leur mise en œuvre.**
- Décision
  - **Faire des tests de montée en charge, sur un clone de notre plateforme ENT**
  - **Identifier le(s) problème(s) (configuration, optimisation, ou architecture)**

La plate-forme ENT dédiée aux tests

- Une architecture identique à la production

  - **Un nouveau frontal web configuré à l'identique**
  - **Un CASv2 dédié aux stress des applications**
  - **10 serveurs d'applications parmi les 20 existants, configurés avec le CAS de stress**
  - **Base de données Apogée de production**

Utilisation des comptes

Contrainte

- **Utilisation des 3500 comptes étudiants pour les besoins du stress dans un cadre légal et sécurisé**

Implémentation

- **Les login avec des mot-de-passe générés sont stockés dans une base de données**
- **Configuration du CAS avec une 2ème méthode d'authentification de type base de données, après LDAP**
- **Filtrage au niveau du CAS pour n'autoriser que les machines de la plateforme de stress**

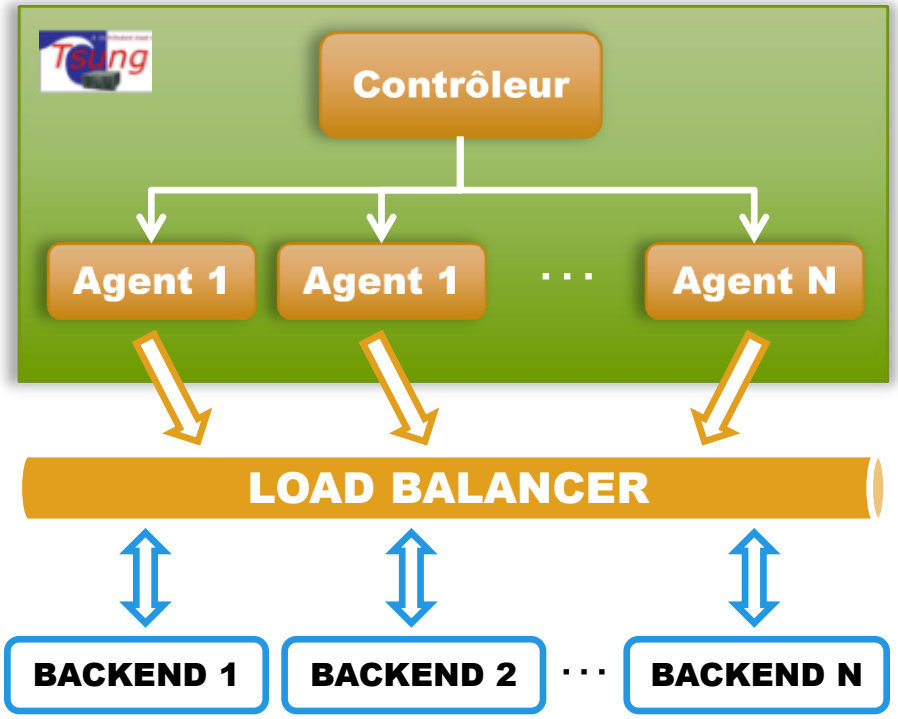
L'outil de stress : TSUNG

Caractéristiques

- **Multi-Protocole : HTTP, LDAP, MYSQL, JABBER…**
- **Performant : écrit en langage objet Erlang**
- **Permet de lancer des tests distribués**
- **Utilise le XML pour la configuration et les scénarios**
- **Simple, pour rendre un scénario dynamique**
- **Plusieurs sessions peuvent être jouer dans le même test de montée en charge**
- **Génère des statistiques, des graphes et des comparatifs**

http://tsung.erlang-projects.org

Plateforme de stress



- Caractérisques

- 15 machines virtuelles, déployées sur des machines affectées à la formation

- Avec la même image :
  - Debian 6, 1Go RAM, 4Go disque
  - Tsung v1.4
  - Un serveur web Apache
  - Plusieurs comptes

- Une authentification vers et entre les VMs par clés SSH

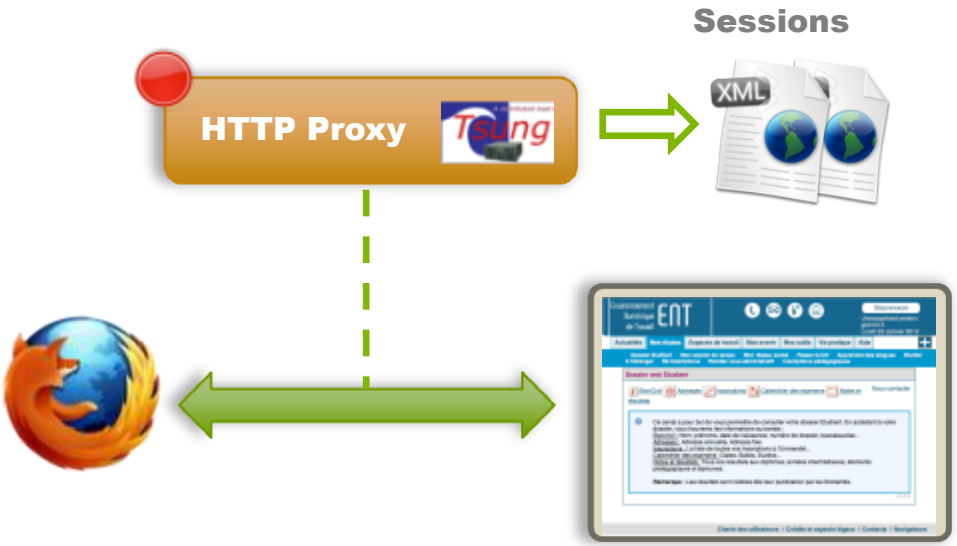
Les scénarios de tests

Profil

- **Utilisent des sessions web issues de la consultation des pages de la portlet esup-mondossierweb**
- **Durée de chaque session varie entre 2 à 5 minutes**
- **Simulent un scénario d'une durée de 5 à 10 minutes avec 3500 à 6000 sessions différentes**
- **Une montée progressive en charge de 10, 15 ou 20 nouvelles sessions par seconde**

Déroulement d'un test de montée en charge

Préparation



- **Créer un scénario via l'enregistreur de sessions**
- **Intégrer le fichier XML du scénario dans le fichier de configuration de Tsung**
- **Modifier le scénario pour le rendre dynamique**
- **Tester le scénario**

Déroulement d'un test de montée en charge

```
<tsung loglevel="notice" version="1.0">
  ....
<options>
   <option name="file_server" id="file_id" value="./userlist.csv"> </option>
</options>
    ...
<sessions>
<session name="rec20110930-1625" probability="100" type="ts_http">
...
<setdynvars sourcetype="file" fileid="file_id" delimiter=";" order="iter">
 <var name="user_name"/> <var name="user_password"/>
</setdynvars>
 ...
<request>
<dyn_variable name="lt"/>
<http url="https://castest.univ-paris5.fr:8450/cas/index.jsp?service=https://ent.univ-
paris5.fr/Login" version="1.1" method="GET"/>
</request>
...
<request subst="true">
<dyn_variable name="urlticket" re="href=&amp;quot;([^&amp;quot;\r\n]*)"/>
<http url="https://castest.univ-paris5.fr:8450/cas/index.jsp?service=https://ent.univ-
paris5.fr/Login" version="1.1" contents="username=%%_user_name%%
&amp;amp;password=%%_user_password%%&amp;amp;lt=%%_lt%%"
content_type="application/x-www-form-urlencoded" method="POST"/>
</request>

<request subst="true"><http url="%%_urlticket%%" version="1.1"
method="GET"/></request>
...
<request>
<dyn_variable name="pltctarget" re="pltc_type_([^\.\r\n]*)"/><http
url="https://ent.univparis5.fr/render.userLayoutRootNode.uP?uP_sparam=activeTab&a
mp;activeTab=3&amp;amp;uP_root=u22l1n6" version="1.1" method="GET"/>
</request>
...

<request subst="true"><http
url="/render.userLayoutRootNode.target.u22l1n6.uP?pltc_target=%%_pltctarget%%.u2
2l1n6&amp;amp;pltc_type_%%_pltctarget%%.u22l1n6=RENDER&amp;amp;pltp_%%_pltctarget
%%.u22l1n6_org.apache.myfaces.portlet.MyFacesGenericPortle
t.VIEW_ID=%2Fstylesheets%2Fetu%2Fnotes.xhtml" version="1.1"
method="GET"/></request>

…

</session>
</sessions>
</tsung>
```

Adaptation

- **URLs de l'authenfication via le CAS**
  - **Login Ticket**
  - **login, et mot-de-passe**
  - **Service Ticket**
- **URLs de la portlet monDossierWeb**
  - **USER_ID associé par uPortal à l'utilisateur**

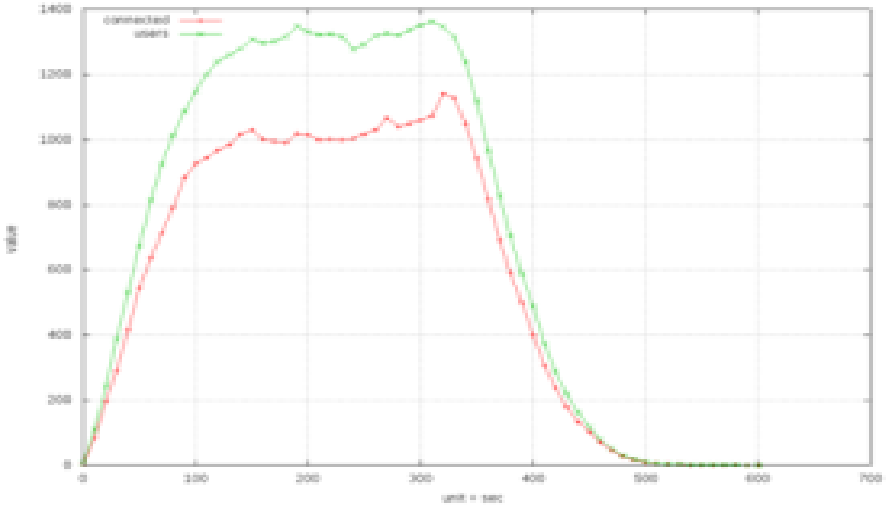
Déroulement d'un test de montée en charge

Exécution

- **Fixer les paramètres du stress**
- **Lancer le stress**
- **Surveiller le comportement de la cible**
- **A la fin du test, générer le rapport et analyser les résultats**
- **Changer les paramètres, redémarrer les Tomcats, et recommencer**

Résultats des tests

Name	highest 10sec mean	lowest 10sec mean	Highest Rate	Mean	Count
connect	0.56 sec	2.34 msec	81.5 / sec	72.61 msec	22643
page	30.15 sec	0.18 sec	231.3 / sec	1.07 sec	64140
request	29.96 sec	14.94 msec	1037.2 / sec	0.24 sec	290520
session	5mn 19sec	48.90 sec	18.2 / sec	1mn 39sec	4424



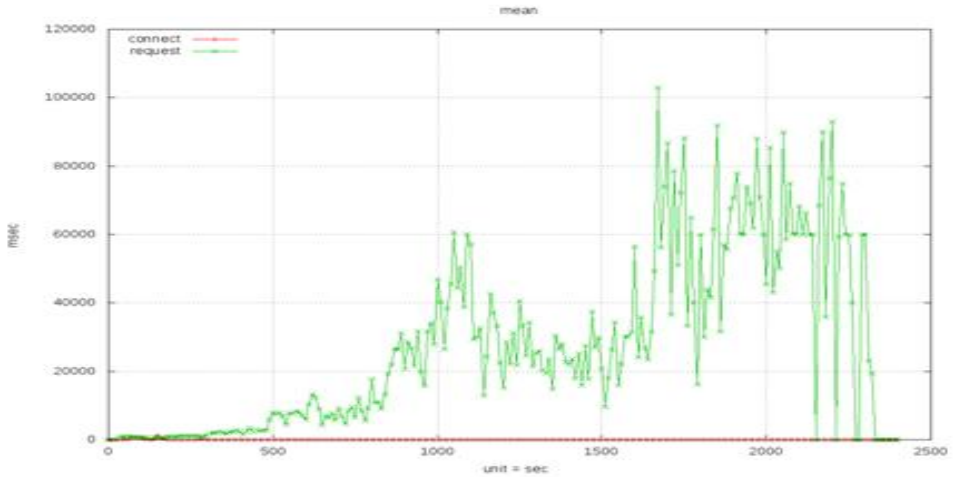
- Scénario 1
  - **Montée en charge, simulant 4500 sessions PAES en 5 minutes, avec 15 nouvelles sessions/s.**
  - **10 backends, avec le *MaxThread* à 500**
- Observations
  - **Temps de réponse satisfaisant (1s)**
  - **Frontal relativement chargé**
  - **Pas d'erreurs HTTP 50X**
  - **Aucun serveur d'application n'est tombé en panne.**

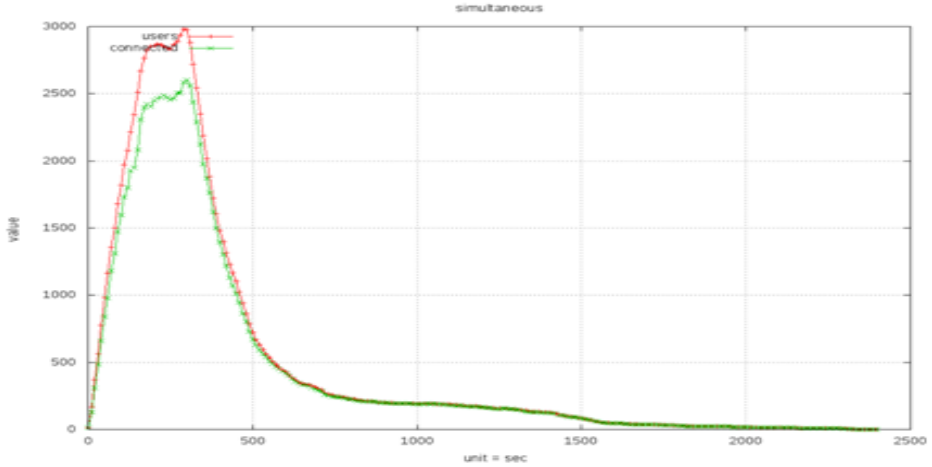
ESUP Portail

UNIVERSITÉ PARIS DESCARTES

Résultats des tests

Name	highest 10sec mean	lowest 10sec mean	Highest Rate	Mean	Count
connect	1.26 sec	2.26 msec	113.9 / sec	0.12 sec	30776
page	4mn 17sec	92.31 msec	301.4 / sec	8.15 sec	86750
request	1mn 43sec	18.14 msec	2313 / sec	1.80 sec	392930
session	35mn 37sec	1mn 0sec	23 / sec	3mn 21sec	5990





Scénario 2

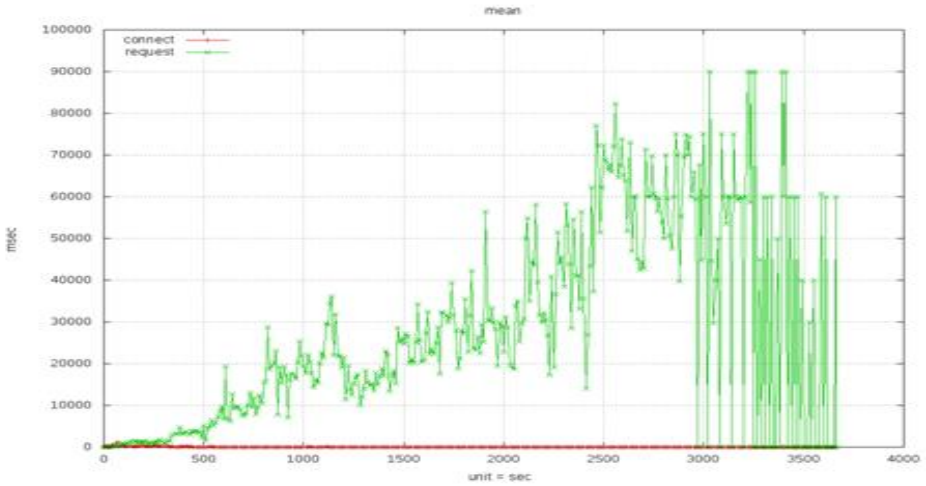
- **Montée en charge, simulant 6000 sessions PAES en 5 minutes, avec 20 nouvelles sessions/s.**
- **10 backends, avec le *MaxThread* à 500**

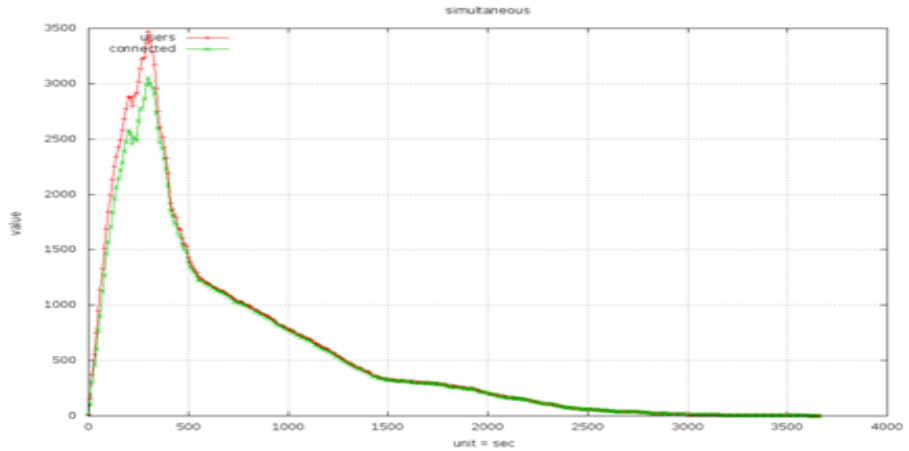
Observations

- **Temps de réponse long (8s)**
- **Frontal est surchargé**
- **Quelques erreurs HTTP 50X**
- **Un seul serveur d'application est tombe en panne.**

Résultats des tests

Name	highest 10sec mean	lowest 10sec mean	Highest Rate	Mean	Count
connect	1.19 sec	2.37 msec	107 / sec	0.25 sec	31480
page	4mn 0sec	0.40 sec	305.1 / sec	18.15 sec	88229
request	1mn 30sec	30.70 msec	1820.2 / sec	4.01 sec	399629
session	56mn 24sec	1mn 0sec	27.2 / sec	5mn 49sec	6016





Scénario 3

- **Montée en charge, simulant 6000 sessions PAES en 5 minutes, avec 20 nouvelles sessions/s.**
- **10 backends, avec le *MaxThread* à 200**

Observations

- **Temps de réponse très long (18s)**
- **Frontal est surchargé**
- **Beaucoup d'erreurs HTTP 50X**
- **Aucun serveur d'application en panne.**

ESUP Portail

UNIVERSITÉ PARIS DESCARTES

Insuffisances techniques et structurelles

- Le frontal web est un goulot d'étranglement
  - **Apache assure les trois fonctions (SSL, répartition de charge et serveur web).**
  - **La surconsommation de la mémoire par Apache lors d'un pic, pénalise toute la plateforme**
- L'absence de redondance fragilise la plateforme
  - **La plateforme est dépendante du bon fonctionnement du frontal**

L'architecture ENT cible

- 4 critères
  - **Haute disponibilité**
  - **Modularité**
  - **Performance**
  - **Simplicité**
- Ingrédients
  - **Hearbeat pour la redondance**
  - **Stunnel pour le SSL**
  - **HAProxy pour la répartition de charge**
  - **Serveur web Nginx**
  - **Un cache pour les pages statiques**
  - **Protocole HTTP remplacera AJP**

Heartbeat

- Pourquoi ?
  - **Permet de gérer la haute disponibilité des ressources dans un cluster en mode Actif/Passif**
  - **Déplacement dynamique/manuel des ressources**
  - **Simple à installer et à configurer**
- Configuration
  - **Gère un cluster de deux répartiteurs en mode Actif/Passif**
  - **Une interface dédiée pour les check Heartbeat**
  - **Deux adresses IP virtuelles, comme ressources**

HAProxy

- Pourquoi ?
  - **Répartiteur stable et très performant, utilisés par des sites à fort trafic**
  - **Gestion dynamique de la disponibilité des backends**
  - **Protège les backends en limitant le nombre de requêtes simultanées**
  - **Simple à installer et à configurer**
- Configuration
  - **En mode HTTP**
  - **Algorithme de répartition de type round-robin**
  - **Persistance des sessions via les cookies**
  - **Disponibilité des backends via des Health-Check**
  - **Limitation des requêtes simultanées par backend**

Nginx

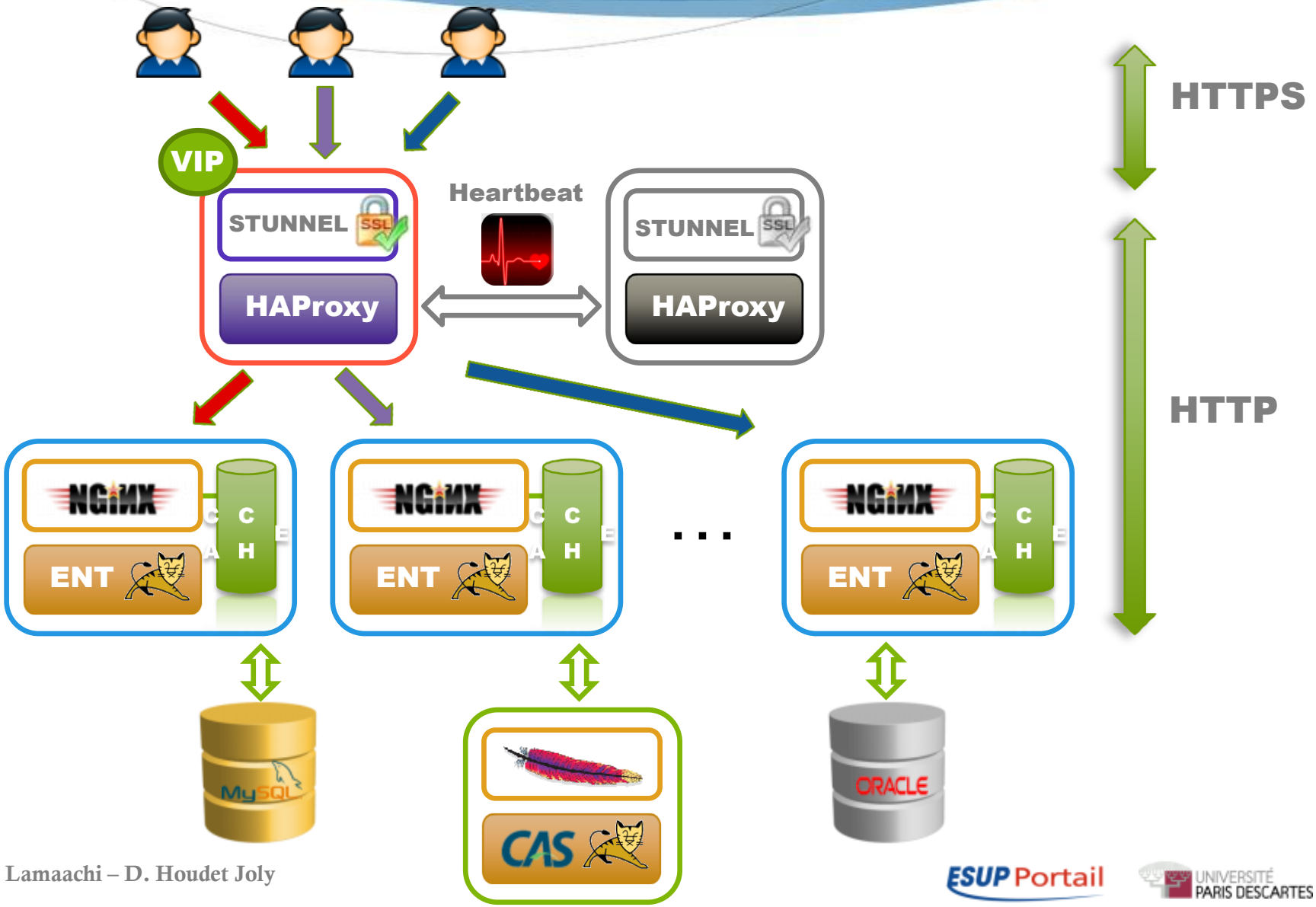
Pourquoi ?

- **Plus performant qu'Apache dans le traitement des pages statiques**
- **Consomme peu de mémoire, même en cas de fort trafic**
- **Simple à installer et à configurer**

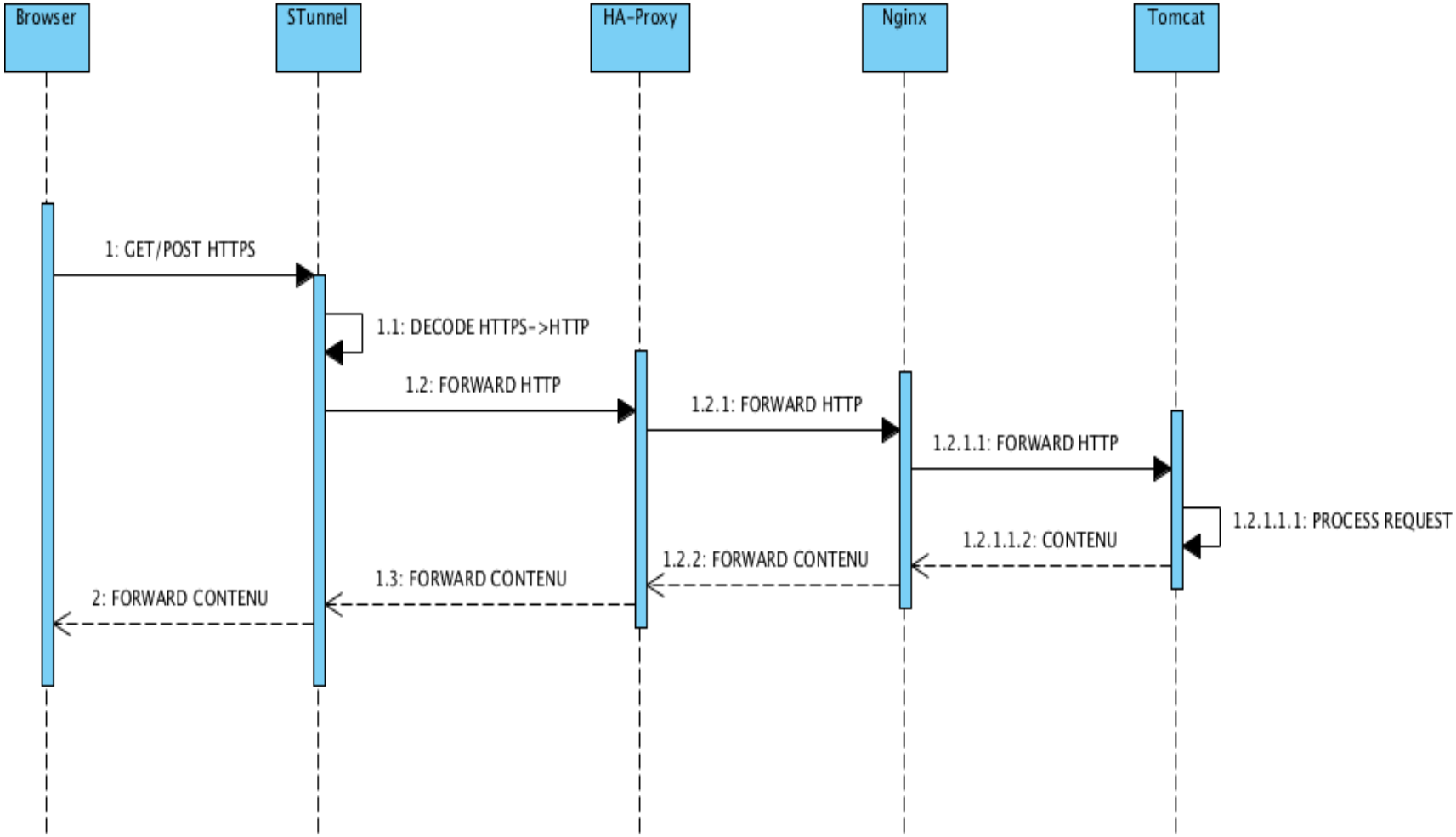
Limites

- **Ne gère pas les fichiers *htaccess***
- **Ne supporte pas *Shibboleth***

Architecture cible de l'ENT



Cheminement d'une requête HTTPS



Adresse IP du client & Tomcat

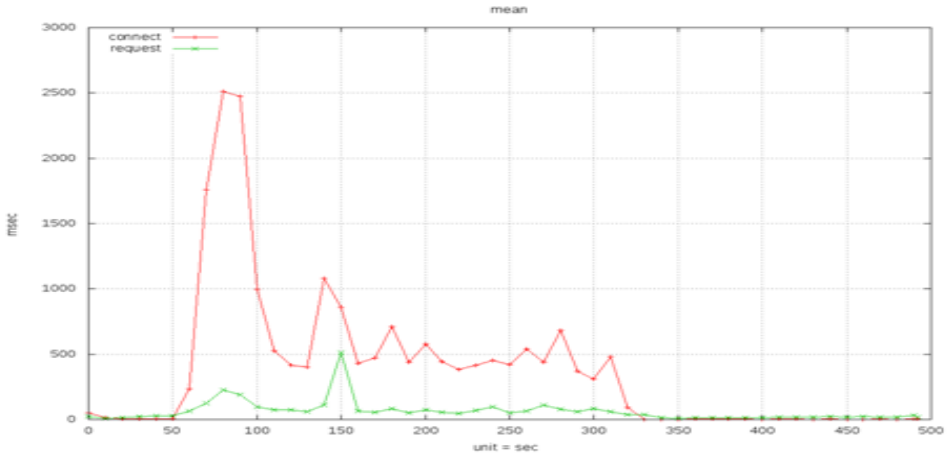
- *Entête HTTP*
  - *Stunnel*
    - Ajoute un champs *X-Forwarded-For* avec l'IP du client
  - *HAProxy*
    - Ajoute un champs *X-Forwarded-Proto* avec la valeur *« https »*

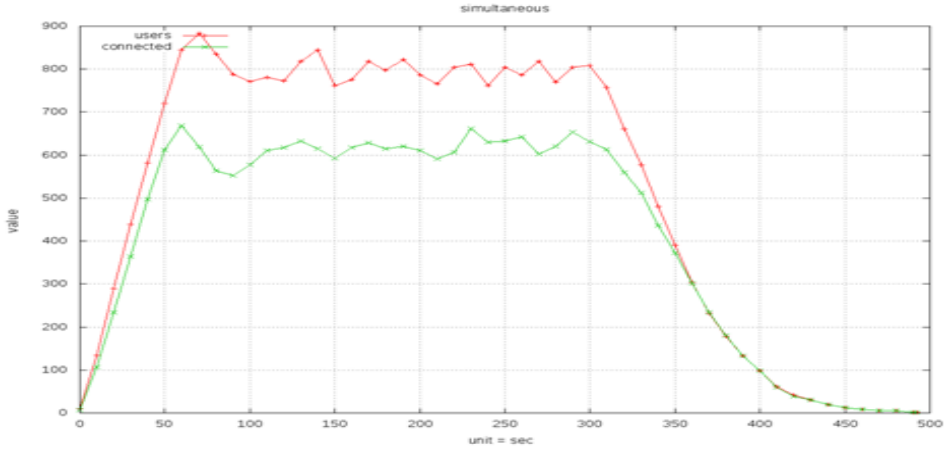
- *Extension RemoteIpValve de Tomcat*
  - *Activation de l'extension* dans le fichier server.xml :

```
<Valve className="org.apache.catalina.valves.RemoteIpValve"
       protocolHeader="X-Forwarded-Proto"
       protocolHeaderHttpsValue="https"
       internalProxies="127\.0\.0\.1" />
```

Les mêmes scénarios sur la plateforme cible

Name	highest 10sec mean	lowest 10sec mean	Highest Rate	Mean	Count
connect	2.11 sec	2.12 msec	50.5 / sec	0.52 sec	12872
page	0.79 sec	56.69 msec	136.2 / sec	0.33 sec	37298
request	0.19 sec	11.34 msec	719.5 / sec	67.65 msec	183935
session	3mn 44sec	26.27 sec	24.3 / sec	56.82 sec	4505





Scénario 1

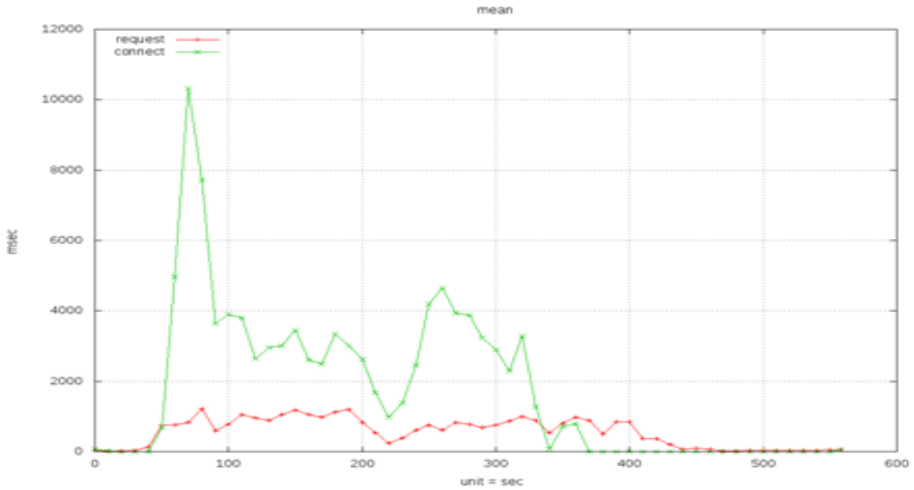
- **Montée en charge, simulant 4500 sessions PAES en 5 minutes, avec 15 nouvelles sessions/s.**
- **6 backends, avec le *MaxThread* à 800**

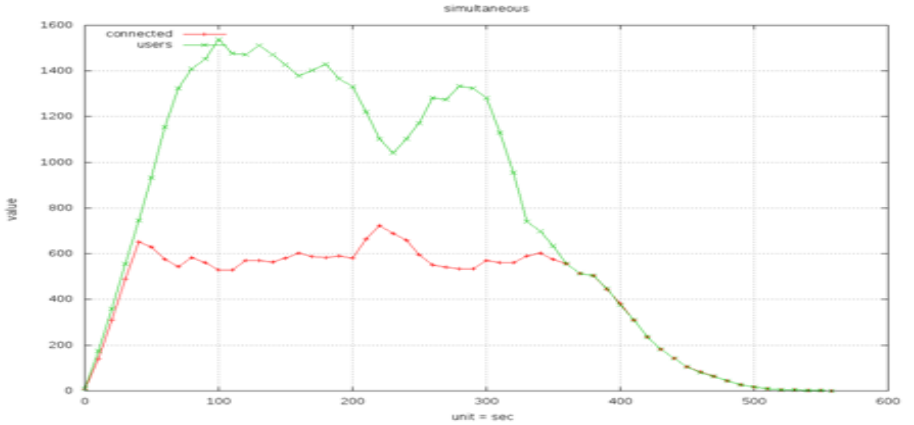
Observations

- **Temps de réponse très satisfaisant (0,3s)**
- **Aucune anomalie, ni surcharge constatée, lors du test**

Les mêmes scénarios sur la plateforme cible

Name	highest 10sec mean	lowest 10sec mean	Highest Rate	Mean	Count
connect	10.33 sec	2.18 msec	57.2 / sec	1.92 sec	9733
page	11.28 sec	56.62 msec	103.5 / sec	4.09 sec	28097
request	1.22 sec	14.31 msec	816.8 / sec	0.66 sec	149490
session	4mn 39sec	11.78 sec	31.8 / sec	1mn 13sec	5938





Scénario 2

- **Montée en charge, simulant 6000 sessions PAES en 5 minutes, avec 20 nouvelles sessions/s.**
- **6 backends, avec le *MaxThread* à 800**

Observations

- **Temps de réponse reste acceptable (4s)**
- **Pas d'erreurs HTTP 50X**

Comparatif des résultats entre les deux plateformes

Plateforme 1 (PFv1)

Name	highest 10sec mean	lowest 10sec mean	Highest Rate	Mean	Count
connect	0.56 sec	2.34 msec	81.5 / sec	72.61 msec	22643
page	30.15 sec	0.18 sec	231.3 / sec	1.07 sec	64140
request	29.96 sec	14.94 msec	1037.2 / sec	0.24 sec	290520
session	5mn 19sec	48.90 sec	18.2 / sec	1mn 39sec	4424

Plateforme 2 (PFv2)

Name	highest 10sec mean	lowest 10sec mean	Highest Rate	Mean	Count
connect	2.11 sec	2.12 msec	50.5 / sec	0.52 sec	12872
page	0.79 sec	56.69 msec	136.2 / sec	0.33 sec	37298
request	0.19 sec	11.34 msec	719.5 / sec	67.65 msec	183935
session	3mn 44sec	26.27 sec	24.3 / sec	56.82 sec	4505

Scénario 1

- **Montée en charge, simulant 4500 sessions PAES en 5 minutes, avec 15 nouvelles sessions/s.**

Observations

- **Temps de réponse PFv2 < 3 X PFv1**
- **Au-delà des 5mn le temps de réponse sur PFv2 reste constant**

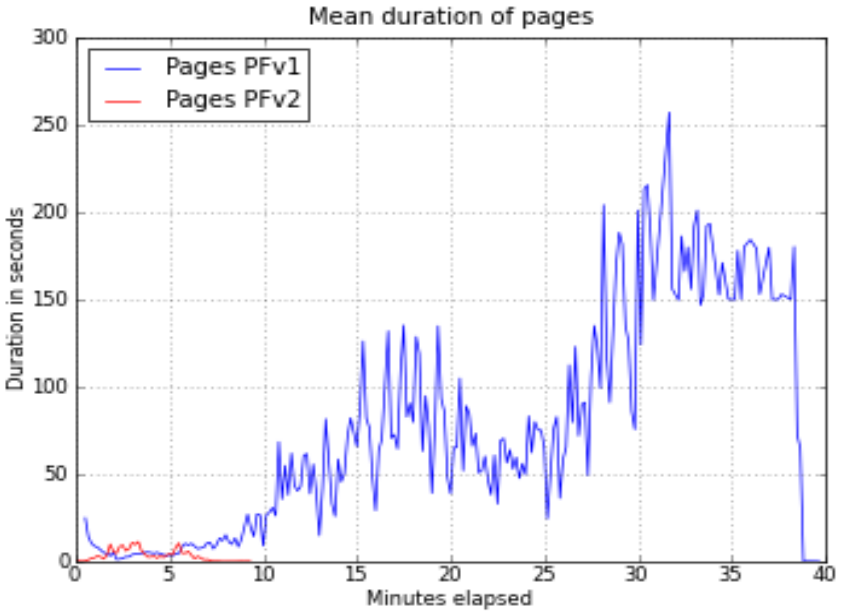
Comparatif des résultats entre les deux plateformes

PFv1

Name	highest 10sec mean	lowest 10sec mean	Highest Rate	Mean	Count
connect	1.26 sec	2.26 msec	113.9 / sec	0.12 sec	30776
page	4mn 17sec	92.31 msec	301.4 / sec	8.15 sec	86750
request	1mn 43sec	18.14 msec	2313 / sec	1.80 sec	392930
session	35mn 37sec	1mn 0sec	23 / sec	3mn 21sec	5990

PFv2

Name	highest 10sec mean	lowest 10sec mean	Highest Rate	Mean	Count
connect	10.33 sec	2.18 msec	57.2 / sec	1.92 sec	9733
page	11.28 sec	56.62 msec	103.5 / sec	4.09 sec	28097
request	1.22 sec	14.31 msec	816.8 / sec	0.66 sec	149490
session	4mn 39sec	11.78 sec	31.8 / sec	1mn 13sec	5938



Scénario 2

- **Montée en charge, simulant 6000 sessions PAES en 5 minutes, avec 20 nouvelles sessions/s.**

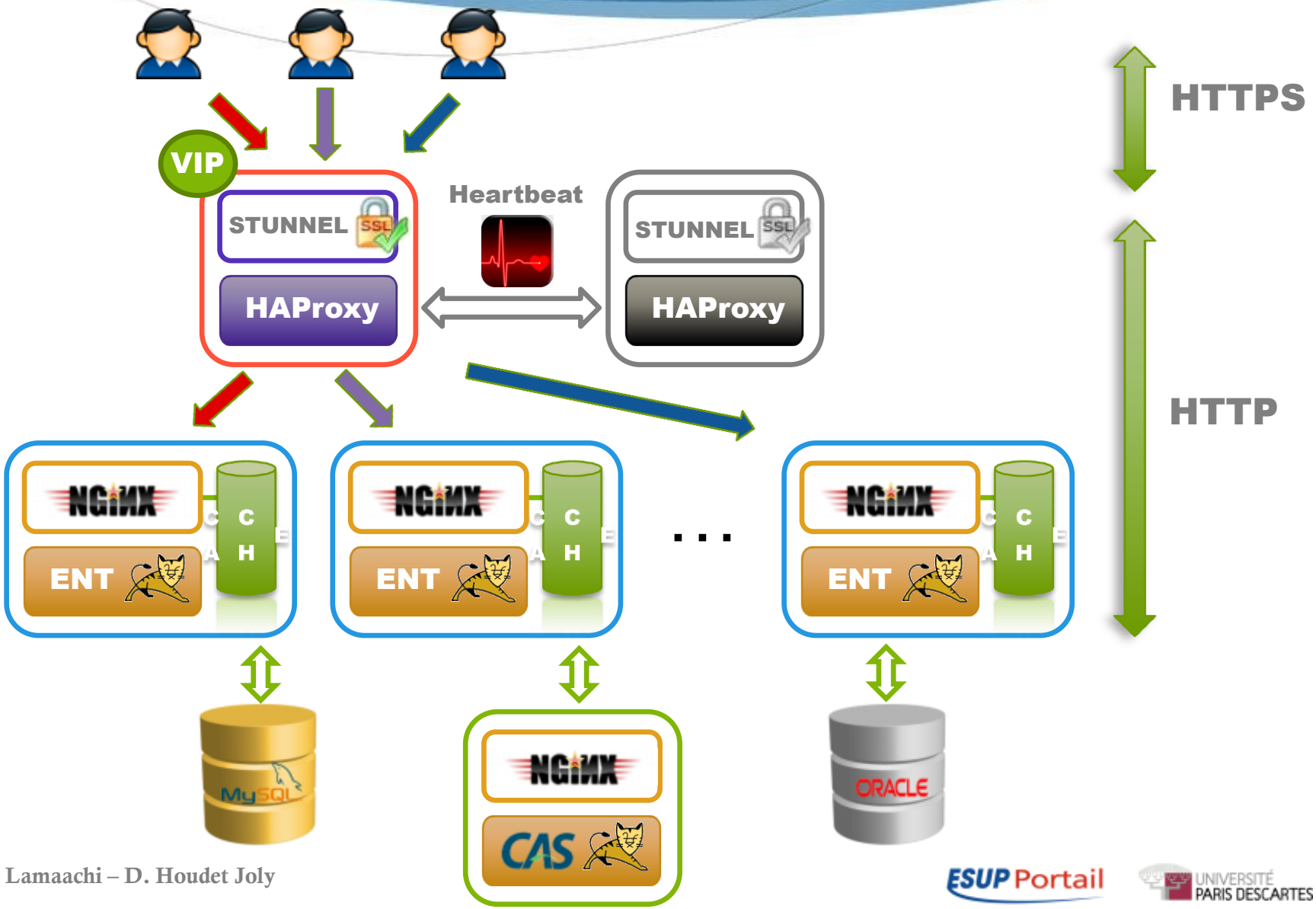
Observations

- **Temps de réponse PFv2 < 2 X PFv1**
- **Au-delà des 5mn le temps de réponse sur PFv2 reste constant**

Et le CAS ?

- Rafraîchi
  - **Un nouvel OS Debian 6, Nginx a remplacé Apache, HTTP a remplacé AJP, et plus de mémoire pour la JVM**
- Adresse IP de l'utilisateur
  - **Nginx ajoute un X-Forwarded-For à l'entête HTTP**
  - **Modification de la classe GenericHandler pour récupérer ce champs**
- Evolutions à venir
  - **Migration en CASv3 sur deux serveurs en répartition de charge**

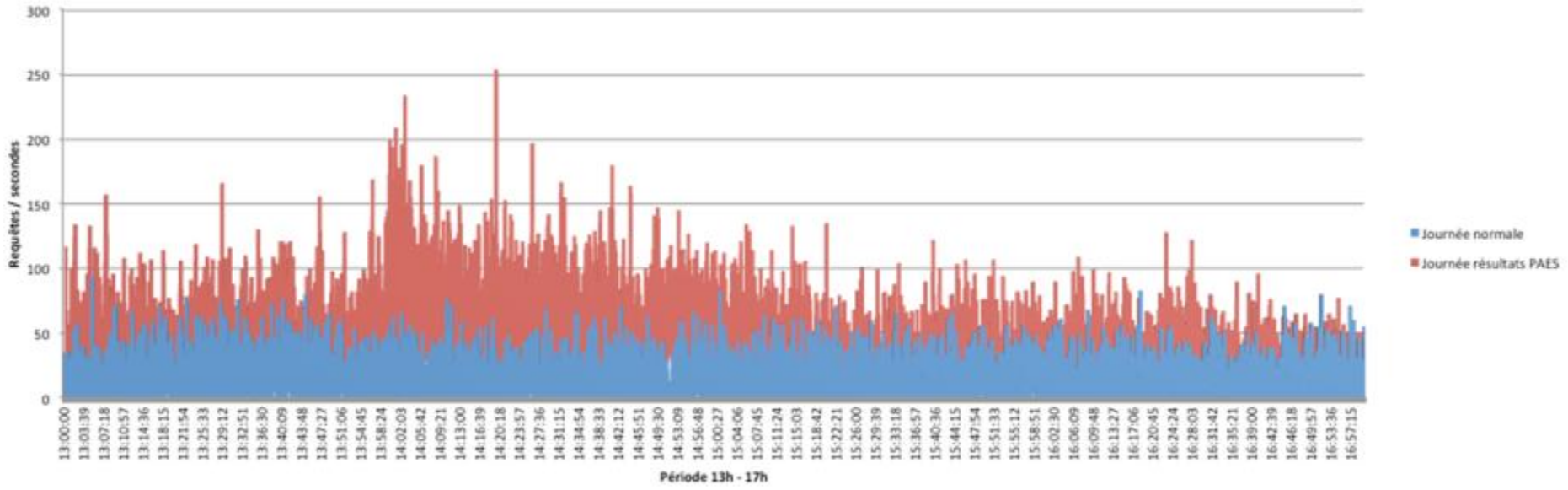
Architecture opérationnelle



Le 27 janvier 2012, jour des résultats PAES…

- Aucune anomalie, ni surcharge constatée
- L'ENT est resté parfaitement opérationnel et réactif
- La plateforme a encaissé le pic de charge
- 1600 nouvelles sessions entre 14H00 et 15H00
- 12000 sessions actives entre 13h et 17h

Comparatif avec une journée normale



Application de la méthode aux services numériques stratégiques

- Une expérience déclinée pour d'autres applications vitales :
  - La plateforme Moodle d'e-learning avec la société Val'Eisti
  - Les lames virtuelles avec la société Tribvun
  - Descartes Media Suite (DMS)
    - Descartes Broadcast
    - Media^2 Descartes

Casting ENT

Direction de l'Informatique et des Systèmes d'Information

Département études et développement :

- **Equipe ENT** — ***Dominique Houdet-Joly, Thierry Sebbar***

Département des moyens informatiques :

- **Benchmarks & HA** — ***Mohamed Lamaachi***
- **Système** — ***Yves Gerday***
- **Base de données** — ***Richard Vatré***

Contact **: mohamed.lamaachi@parisdescartes.fr**

Crédits graphisme
Eddy Pelaic

ESUP Portail

UNIVERSITÉ PARIS DESCARTES