



Geolocalized web usage mining

Sommaire



INTRODUCTION



ARCHITECTURE DU
PROJET



ANALYSES ET
PRÉSENTATION DES
LOGS



AUTHENTIFICATION



CLOUD
COMPUTING
PROJECT



VISUALISATION DU
PROJET



PROBLÈMES
RENCONTRÉS



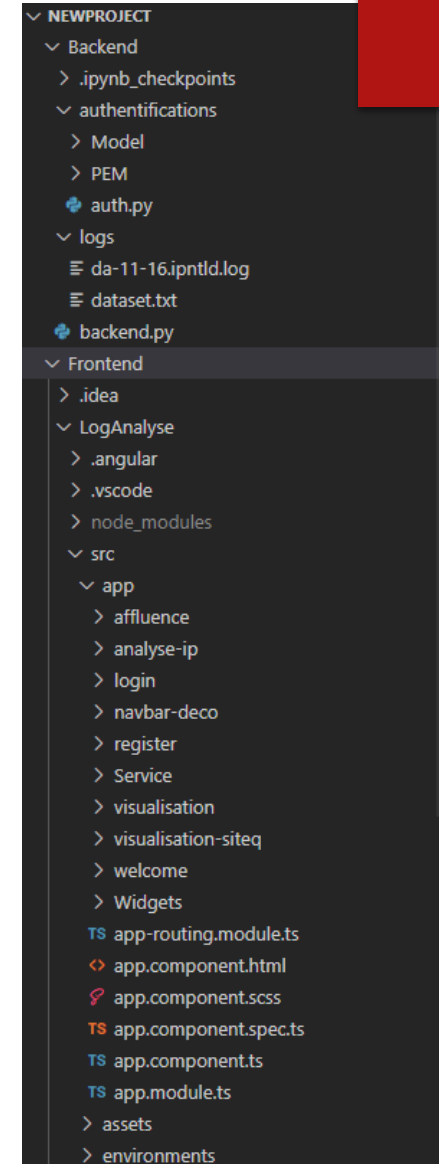
CONCLUSION

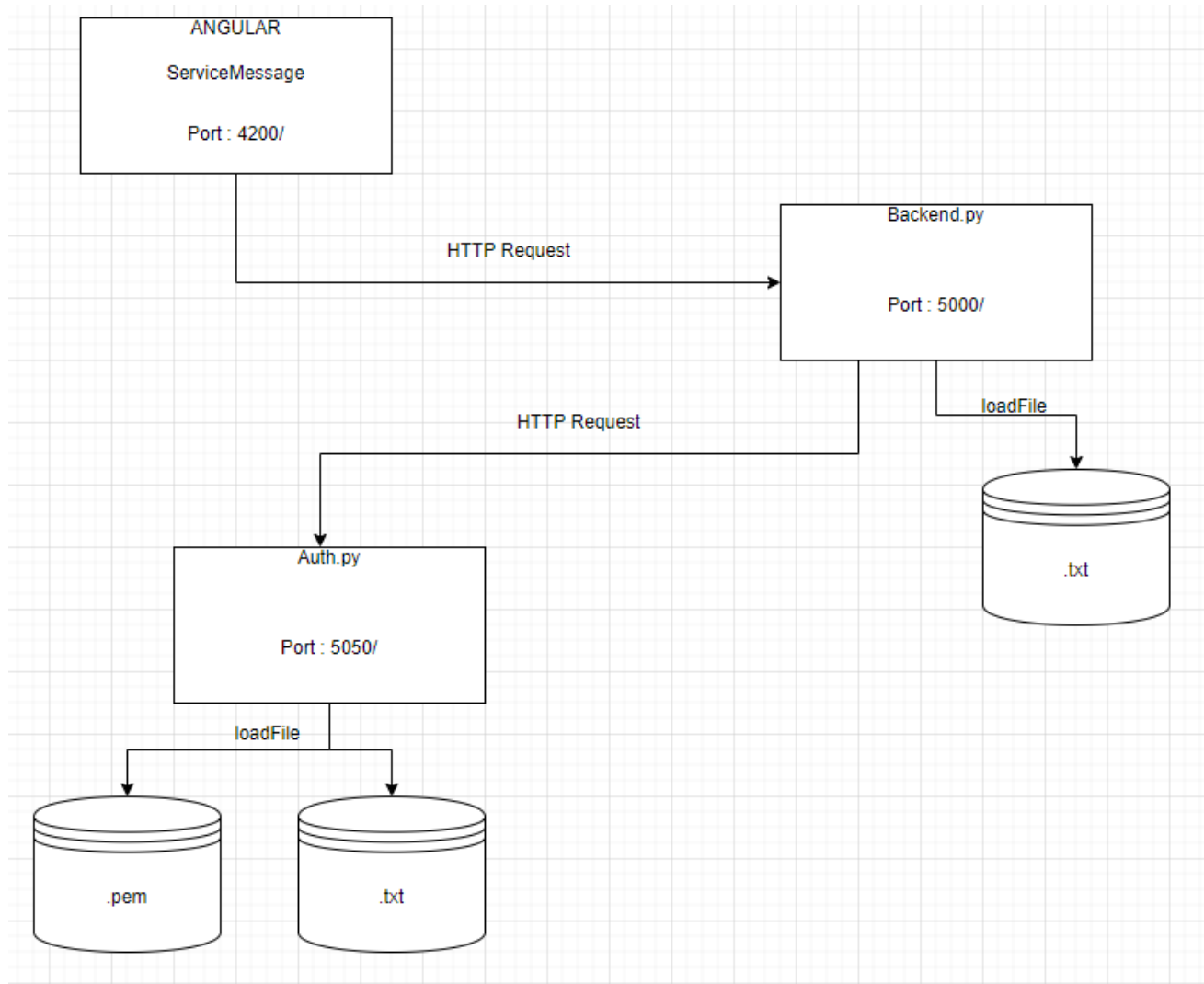
Introduction

- ▶ Objectifs :
 - ▶ Se servir des fichiers logs fournis par le serveur afin de faire des analyses sur la navigation des utilisateurs.
 - ▶ Créer une interface web permettant de visualiser les différentes analyses.
 - ▶ Proof of concept
 - ▶ Organisation

Architecture du projet

- ▶ Mise en place d'une architecture à 3 niveaux
 - ▶ Frontend avec l'Angular :
 - ▶ Backend avec Flask :
 - ▶ BD : CSV
- ▶ Architecture MVC
 - ▶ View implanter sur le Front
 - ▶ Contrôleur
- ▶ Communication entre les Parties
 - ▶ Service Messages
 - ▶ Fonction Filtrage DataFrame





Architecture du projet

Analyses et présentation des logs

► Fichier formaté à générer

```
ip1664.com - - [16/Nov/2005:00:00:43 -0500] "GET /robots.txt HTTP/1.0" 200 173 "-" "msnbot/1.0 (+http://search.msn.com/msnbot.htm)"
ip1664.com - - [16/Nov/2005:00:00:43 -0500] "GET /gpspubs/sigkdd-kdd99-panel.html HTTP/1.0" 200 14199 "-" "msnbot/1.0 (+http://search.msn.com/msnbot.htm)"
ip1115.unr - - [16/Nov/2005:00:01:00 -0500] "GET /news/99/n23/i12.html HTTP/1.1" 200 3171 "http://discount-blah1.professional-doctor.com/" "Mozilla/4.0 (co
ip2283.unr - - [16/Nov/2005:00:01:02 -0500] "GET /dmcourse/data_mining_course/assignments/assignment-3.html HTTP/1.1" 200 8090 "http://www.google.com/searc
ip2283.unr - - [16/Nov/2005:00:01:03 -0500] "GET /dmcourse/dm.css HTTP/1.1" 200 155 "http://www.kdnuggets.com/dmcourse/data_mining_course/assignments/assig
ip1389.net - - [16/Nov/2005:00:02:46 -0500] "GET /gpspubs/kdd99-est-ben-lift/sld021.htm HTTP/1.1" 200 1385 "http://www.google.com/search?hs=JnE&hl=en&lr=&c
ip1389.net - - [16/Nov/2005:00:02:46 -0500] "GET /gpspubs/kdd99-est-ben-lift/img021.gif HTTP/1.1" 200 7465 "http://www.kdnuggets.com/gpspubs/kdd99-est-ben-
ip1389.net - - [16/Nov/2005:00:02:47 -0500] "GET /favicon.ico HTTP/1.1" 200 899 "http://www.kdnuggets.com/gpspubs/kdd99-est-ben-lift/sld021.htm" "Mozilla/4
ip1946.com - - [16/Nov/2005:00:02:49 -0500] "GET /news/2001/n10/15i.html HTTP/1.0" 200 4214 "-" "Mozilla/5.0 (compatible; Yahoo! Slurp; http://help.yahoo.c
ip992.unr - - [16/Nov/2005:00:02:52 -0500] "GET /aps/bt4-a.sol_crm.re.html HTTP/1.0" 304 - "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT; MS Search 4.
ip2213.net - - [16/Nov/2005:00:04:12 -0500] "GET / HTTP/1.1" 200 12453 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1)"
ip2213.net - - [16/Nov/2005:00:04:16 -0500] "GET /kdr.css HTTP/1.1" 304 - "http://www.kdnuggets.com/" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; S
ip2213.net - - [16/Nov/2005:00:04:16 -0500] "GET /images/KDnuggets_logo.gif HTTP/1.1" 304 - "http://www.kdnuggets.com/" "Mozilla/4.0 (compatible; MSIE 6.0;
ip2213.net - - [16/Nov/2005:00:04:16 -0500] "GET /images/kdnuggets.co.jp.gif HTTP/1.1" 304 - "http://www.kdnuggets.com/" "Mozilla/4.0 (compatible; MSIE 6.0
ip2213.net - - [16/Nov/2005:00:04:16 -0500] "GET /aps/aw1.js HTTP/1.1" 304 - "http://www.kdnuggets.com/" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1
ip2213.net - - [16/Nov/2005:00:04:16 -0500] "GET /images/newy.gif HTTP/1.1" 304 - "http://www.kdnuggets.com/" "Mozilla/4.0 (compatible; MSIE 6.0; Windows N
ip2213.net - - [16/Nov/2005:00:04:16 -0500] "GET /aps/bt4-a.ind.gif HTTP/1.1" 304 - "http://www.kdnuggets.com/" "Mozilla/4.0 (compatible; MSIE 6.0; Windows
ip2213.net - - [16/Nov/2005:00:04:16 -0500] "GET /aps/f-spss-h2.ind.gif HTTP/1.1" 304 - "http://www.kdnuggets.com/" "Mozilla/4.0 (compatible; MSIE 6.0; Win
ip2213.net - - [16/Nov/2005:00:04:16 -0500] "GET /aps/t-salf-sd12.c12.gif HTTP/1.1" 200 28266 "http://www.kdnuggets.com/" "Mozilla/4.0 (compatible; MSIE 6.
ip2213.net - - [16/Nov/2005:00:04:16 -0500] "GET /aps/f-ang-dm21.ind.gif HTTP/1.1" 304 - "http://www.kdnuggets.com/" "Mozilla/4.0 (compatible; MSIE 6.0; Wi
ip2213.net - - [16/Nov/2005:00:04:17 -0500] "GET /aps/rw1.js HTTP/1.1" 304 - "http://www.kdnuggets.com/" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1
ip2213.net - - [16/Nov/2005:00:04:17 -0500] "GET /aps/x-salf-tdm.c15.gif HTTP/1.1" 200 92 "http://www.kdnuggets.com/" "Mozilla/4.0 (compatible; MSIE 6.0; W
ip2213.net - - [16/Nov/2005:00:04:17 -0500] "GET /aps/r-pur-1.c2.gif HTTP/1.1" 304 - "http://www.kdnuggets.com/" "Mozilla/4.0 (compatible; MSIE 6.0; Window
ip2213.net - - [16/Nov/2005:00:04:17 -0500] "GET /swh.gif?width=1280&height=768 HTTP/1.1" 304 - "http://www.kdnuggets.com/" "Mozilla/4.0 (compatible; MSIE
ip1253.com - - [16/Nov/2005:00:04:25 -0500] "GET /news/2004/n24/18i.html HTTP/1.0" 200 2757 "-" "psbot/0.1 (+http://www.picsearch.com/bot.html)"
ip2213.net - - [16/Nov/2005:00:04:47 -0500] "GET /aps/r-pur-1.c2.re.html HTTP/1.1" 200 265 "http://www.kdnuggets.com/" "Mozilla/4.0 (compatible; MSIE 6.0;
ip1886.com - - [16/Nov/2005:00:05:24 -0500] "GET /news/2001/n21/28i.html HTTP/1.0" 200 4542 "-" "Mozilla/5.0 (compatible; Yahoo! Slurp; http://help.yahoo.c
ip1660.unr - - [16/Nov/2005:00:05:47 -0500] "HEAD / HTTP/1.1" 200 0 "-" "InternetSeer.com"
```

Analyses et présentation des logs

- CSV → Date / Heure / Page Consultée / IP / Site Visité / StatusCode / Octets transférés

```
2015-11-16 22:34:36 usep-pinchinades 68.180.229.29 associations.paysdaixassociations.org 301 29828
2015-11-16 22:34:49 usep-pinchinades 68.180.229.29 associations.paysdaixassociations.org 500 12562
2015-11-16 22:35:03 association-regionale-pour-le-developpement-de-l-enseigne-3 68.180.228.251 agneaux-50.ville.mygaloo.fr 301 171
2015-11-16 22:35:05 association-regionale-pour-le-developpement-de-l-enseigne-3 68.180.228.251 agneaux-50.ville.mygaloo.fr 500 1968
2015-11-16 22:35:14 ascair 68.180.230.237 paris-75.ville.mygaloo.fr 301 156
2015-11-16 22:35:16 ascair 68.180.230.237 paris-75.ville.mygaloo.fr 500 1968
2015-11-16 22:35:29 arc-animation-entrainement-a2e-2 208.115.111.66 associations.paysdaixassociations.org 301 171
2015-11-16 22:35:31 aclap-action-de-coordination-de-lieux-et-d-accueil-aux-pers 208.115.111.66 associations.paysdaixassociations.org 301 187
2015-11-16 22:35:33 renesens-2 208.115.111.66 associations.paysdaixassociations.org 301 171
2015-11-16 22:37:40 les-archers-de-l-abbaye-2 82.227.35.120 massy-91.ville.mygaloo.fr 301 46
2015-11-16 22:37:42 les-archers-de-l-abbaye-2 82.227.35.120 massy-91.ville.mygaloo.fr 500 1218
2015-11-16 22:37:46 tja-kon-fesi-avenir-2 68.180.230.188 maripasoula-973.ville.mygaloo.fr 301 171
2015-11-16 22:37:46 association-sport-et-musique-de-bras-fusil-2 146.185.234.48 saint-benoit-974.ville.mygaloo.fr 301 109
2015-11-16 22:37:48 tja-kon-fesi-avenir-2 68.180.229.161 maripasoula-973.ville.mygaloo.fr 500 2000
2015-11-16 22:37:52 association-sport-et-musique-de-bras-fusil-2 146.185.234.48 saint-benoit-974.ville.mygaloo.fr 500 2750
2015-11-16 22:38:09 union-economique-commerciale-de-l-agglomeration-rochelaise- 146.185.234.48 la-rochelle-17.ville.mygaloo.fr 301 46
2015-11-16 22:38:10 union-economique-commerciale-de-l-agglomeration-rochelaise- 146.185.234.48 la-rochelle-17.ville.mygaloo.fr 500 1312
2015-11-16 22:38:39 les-films-du-minotaure-2 146.185.234.48 le-creusot-71.ville.mygaloo.fr 301 93
2015-11-16 22:38:43 les-films-du-minotaure-2 146.185.234.48 le-creusot-71.ville.mygaloo.fr 500 3671
2015-11-16 22:38:49 alizee-club 68.180.230.188 marseille-13.ville.mygaloo.fr 301 171
2015-11-16 22:39:00 alizee-club 68.180.230.188 marseille-13.ville.mygaloo.fr 500 2062
2015-11-16 22:40:50 association-des-commerçants-de-ville-port-2 146.185.234.48 saint-nazaire-44.ville.mygaloo.fr 301 78
2015-11-16 22:40:59 yogaroma-2 68.180.228.96 bonzac-33.ville.mygaloo.fr 301 171
2015-11-16 22:41:02 association-des-commerçants-de-ville-port-2 146.185.234.48 saint-nazaire-44.ville.mygaloo.fr 500 1921
2015-11-16 22:41:04 yogaroma-2 68.180.228.96 bonzac-33.ville.mygaloo.fr 500 1828
2015-11-16 22:41:48 ecole-de-secourisme-lansarguoise-2 37.187.169.135 lansargues-34.ville.mygaloo.fr 301 156
2015-11-16 22:41:53 ecole-de-secourisme-lansarguoise-2 37.187.169.135 lansargues-34.ville.mygaloo.fr 500 1140
2015-11-16 22:42:03 ecole-de-secourisme-lansarguoise-2 37.187.169.135 lansargues-34.ville.mygaloo.fr 500 1125
2015-11-16 22:42:08 ecole-de-secourisme-lansarguoise-2 37.187.169.135 lansargues-34.ville.mygaloo.fr 500 1125
2015-11-16 22:42:20 ecole-de-secourisme-lansarguoise-2 37.187.169.135 lansargues-34.ville.mygaloo.fr 500 1078
```

Analyses et présentation des logs

Fichier formaté à générer



CSV → Date / Heure / Page Consultée / IP /
Site Visité / StatusCode / Octets transférés

Analyses et présentation des logs

- ▶ Affluence sur un site donné
- ▶ Pages les plus consultées d'un site
- ▶ Analyse de navigation d'une IP (Affluence et fréquence d'accès sur un site)
- ▶ Géolocalisation des connexions pour un site donné



Authentification

- ▶ API REST pour l'authentification
 - ▶ Cryptage des données grâce à **RSA**
 - ▶ isAuthenticate lancé à chaque fonctions du BackEnd
 - ▶ Service d'authentification sur le FrontEnd
-
- ⚠ Cookie de login utilisé pour assuré l'authentification (pas de code crypté)
 - ⚠ Mots de passe passé en dur durant le POST



Cloud Computing Project

- ▶ Parallèle avec le projet Cloud Computing
- Dockeriser notre application :
 - 3 Dockerfile → Un pour l'application Angular et deux pour les micro services (API)
 - Docker compose
- ▶ Publication sur AWS à l'aide de l'ECS (Elastic Container Service)




Visualisation du projet

LogAnalyse

localhost:4200/login

LogAnalyse Home Visualisation Visualisation site Cartographie IP Analyse Graphiques Deconnexion



Email address

Password

☒ Remember me

[Forgot password?](#)

Sign in

Register

Problèmes rencontrés

- ▶ Comment gérer des requêtes avec une quantité de donnée assez élevée ?
- ▶ Comment accélérer les processus de chargement ?
- ▶ Choisir le graphique le plus pertinent pour représenter l'analyse
- ▶ Manipulation de GCP
- ▶ Manipulation de Github



Conclusion

▶ Améliorations possibles

- ▶ Interface IHM
- ▶ Introduire des notions de machine Learning
- ▶ Augmenter le nombre de thread pour permettre un meilleur délai
- ▶ Utiliser une base de donnée comme MongoDB

▶ Les Compétences Acquises :

▶ Compétences Techniques

- Prise en main du logiciel Pycharm ou Webstorm
- Manipulation d'API
- Réalisation complète d'une Application Web

▶ Compétences Organisationnel

- Réunion avec des objectifs précis
- Mise en place d'un Plan d'action
- Gestion du temps

Merci pour votre Attention

MEYA George

BARAR Bilal

MONIMEAU Paul

