

# INTRODUCTION AU SYSTEME D'EXPLOITATION

## Sujet SAE 1.03

P. Mely - IUT de Metz – R1.04 - Année 2022-2023

**Conditions** : à faire en binôme, soutenance à partir de la semaine du 16/01/22 : le planning sera donné ultérieurement.

**But** : Travailler à partir de fichier de logs (journalisation) d'un serveur Apache pour sortir des statistiques sur les connexions et les erreurs.

### Contexte :

**Serveur Web et Apache** : un serveur Web permet de donner une représentation à des données de contenus Web. Dans ce but le programme communique avec un client Web (navigateur). Le protocole réseau utilisé est le protocole http (ou https). En général, le navigateur envoie une requête au serveur Web et ce dernier lui répond sous forme de page html.

Le serveur Web http Apache (plus communément appelé serveur Apache) est un des serveurs Web les plus utilisés. Il existe aussi Nginx, Tomcat, Microsoft IIS, etc.

Pour administrer un serveur web, il est nécessaire de disposer d'un retour d'informations à propos de l'activité et des performances du serveur, ainsi que toute panne/problème qui pourrait survenir. Le serveur Apache installe des fichiers qui sont alimentés régulièrement pour y parvenir. C'est ce qu'on appelle la journalisation. Nous allons étudier le fichier access.log et error.log dont vous avez un exemple sur Arche en annexe.

Le fichier **access.log** enregistre toutes les requêtes que traite ce dernier. Ce fichier permet généralement d'avoir des informations qui vont servir à réaliser des statistiques. Le but de cette SAE est justement de mettre en forme ces statistiques. Le fichier affiche :

- L'adresse IP du client (hôte distant) qui a envoyé une requête au serveur.
- L'utilisateur (entre tiret) qui a envoyé la requête.
- La date et l'heure de la requête
- La ligne de la requête du client : la méthode utilisée par le client, la ressource demandée, le protocole (placée entre les guillemets)
- Le code statut que le serveur retourne au client
- La taille de l'objet retourné au client

Le fichier **error.log** (journal des erreurs) permet de connaître des informations de diagnostic et d'enregistrer toutes les erreurs qui surviennent lors du traitement des requêtes. C'est le fichier que regarde l'administrateur du serveur s'il y a un problème. Il reprend plusieurs informations :

- Le premier champ de l'entrée du fichier correspond à la date et l'heure du message.
- Le second champ reprend le type d'erreur
- Le troisième correspond aux process
- Le quatrième, l'adresse IP et le port réseau (délimité par : )
- Enfin la dernière partie affiche le message d'erreur

Le stockage d'informations dans ces journaux n'est que le point de départ de la gestion de la journalisation. L'étape suivante consiste à analyser ces informations de façon à pouvoir en extraire des statistiques utiles. **C'est le but de cette SAE.**

Au préalable, on dépose les fichiers sur un répertoire, qu'on appellera base pour la suite de l'explication.

### **1) 1er script : archivage des fichiers acces.log et error.log**

Le script aura comme paramètre le répertoire base.

Ce script va permettre de copier les fichiers access.log et error.log pour les archiver en fonction de la date.

On regarde dans le répertoire base et pour chaque fichier qui se trouve dans le répertoire, on le copie dans un répertoire en fonction de l'année et du mois du jour, répertoire qu'il faudra créer s'il n'existe pas en gérant les messages d'erreur correspondants.

Les fichiers seront alors préfixés par le jour (JJ\_). On aura, par exemple dans le répertoire /base/2022/02, les fichiers 01\_access.log, 02\_access.log, etc 01\_error.log, 02\_error.log, etc. Pour chaque fichier copié, on vide (remise à zéro) le fichier initial dans le répertoire base.

### **2) 2ème script : Statistiques sur le fichier access.log par jour**

Le script aura deux paramètres : le répertoire base et la date du jour pour laquelle on veut des statistiques. On fera les traitements sur les fichiers access.log en fonction de la date saisie en paramètre.

Ce script se présentera sous la forme d'un menu qui va permettre d'avoir plusieurs statistiques sur le jour saisi en paramètre :

- Après avoir saisi une adresse IP, on aura le nombre de requêtes différentes effectuées par cette adresse.
- Calculer le nombre de requêtes différentes, par heure de la journée.
- Après avoir saisi un nom d'utilisateur, calculer le nombre de code statut différent par utilisateur.
- Après avoir saisi un nom d'utilisateur, afficher toutes les IP différentes de l'utilisateur.

### **3) 3ème script : Statistiques sur le fichier access.log par mois**

On effectuera les mêmes statistiques que le 2<sup>ème</sup> script mais sur un mois (avec un affichage adapté au mois) que l'on saisira. Les deux scripts pourront être réunis en un seul et faire l'objet d'un paramétrage différent en fonction du jour ou du mois.

### **4) 4ème script : Statistiques sur le fichier error.log par jour**

Le script aura deux paramètres : le répertoire base et la date du jour pour laquelle on veut des statistiques. On fera les traitements sur les fichiers error.log en fonction de la date saisie en paramètre.

Ce script se présentera sous la forme d'un menu qui va permettre d'avoir plusieurs statistiques sur le jour saisi en paramètre :

- Pour une journée, afficher la totalité des erreurs en mettant en forme les données :
  - Afficher la date en français
  - Enlever les caractères spéciaux (type [, ], :)
  - Mettre en forme pour que les champs soient bien en colonne
- Même question que précédemment mais la sortie sera un fichier dont l'extension sera en .imp, sur un répertoire *sortie* (que l'on créera s'il n'existe pas)
- Le nombre d'erreurs sur une journée
- Les nombres de type d'erreurs différents correspondant dans une journée
- En saisissant une adresse IP, avoir les erreurs en totalité de cette adresse IP
- En saisissant un pid (process), avoir toutes les erreurs correspondantes
- En saisissant un type d'erreur, afficher les différents messages d'erreur correspondants

#### 5) 5ème script : Statistiques sur le fichier error.log par mois

On effectuera les mêmes statistiques que le 2<sup>ème</sup> script mais sur un mois (avec un affichage adapté au mois) que l'on saisira. Les deux scripts pourront être réunis en un seul et faire l'objet d'un paramétrage différent en fonction du jour ou du mois.

**Rendu de vos travaux** : Vous ferez un rapport pour expliquer vos choix. Vous expliquerez qui à fait quoi au sein de votre binôme. Sur Arche, vous rendrez le ou les scripts et votre rapport avant le 12/01, 18H en zippant l'ensemble de votre projet.

Source : <https://httpd.apache.org>