
Enquête au coeur d'un "service"

Retrouver qui fait quoi pourquoi comment où
et quand...

150+

log samples

Différences et similitudes

format ? contenus ?

But :

- Reconnaître le format
- Comprendre le contenu, repérer les infos
- Deviner les erreurs

Trouvez les traces

Pourquoi ?

- qui ? / fait quoi ? / à quel moment ? / par quel moyens ? / avec quel résultats ?
- Pourquoi cette erreur ?

Il y a des obligations légales à garder des traces et à les effacer suivant le contexte (RGPD, utilisation en interne...).





1. Fichiers

Lisible, en clair ou cryptés les fichiers de log contiennent des informations que les applications/services veulent bien écrire.

- **Dépendant du service**
Chaque application à son propre format
- **Base commune**
Un base de format avec la date / la sévérité / un message
- **Verbeux**
Le contenu est très vite important. Une compression est obligatoire et un vidage régulier est nécessaire.

Une ligne par seconde :

Taille en fin de journée ?



Taille

Une ligne contient :

- une date
- une heure ou un timestamp
- une sévérité
- une texte libre

Solutions :

- Compression
- Filtrage
- Archivage
- Suppression

(Et la réduction des entrées)



Liens

Pour la compression,
regardez [ici](#)

Un serveur écrit dans
DES DIZAINES

de fichiers de logs.

Un service peut écrire
dans plusieurs fichiers de
logs.

Plusieurs services
peuvent écrire dans un
même fichiers

rw-r--r--	1 root	adm	455	Dec 20	03:57	apport.log.1
rw-r--r--	2 root	root	4.0K	Dec 24	02:29	apt
rw-r--r--	1 syslog	adm	424K	Dec 24	02:45	auth.log
rw-r--r--	1 root	root	5.3K	Dec 23	22:08	boot.log
rw-r--r--	1 root	root	61K	Aug 15	2015	bootstrap.log
rw-r--r--	1 root	utmp	1.9K	Dec 16	12:36	btmpt
rw-r--r--	1 root	utmp	6.0K	Nov 30	10:36	btmpt.1
rw-r--r--	2 root	root	4.0K	Dec 24	02:29	cups
rw-r--r--	3 root	root	4.0K	Dec 24	02:29	dist-upgrade
rw-r--r--	1 root	adm	171K	Aug 7	08:34	dmesg
rw-r--r--	1 root	adm	119K	Aug 7	08:15	dmesg.0
rw-r--r--	1 root	root	71K	Dec 24	02:38	dpkg.log
rw-r--r--	1 root	root	3.0K	Nov 8	02:40	dpkg.log.1
rw-r--r--	1 root	root	32K	Dec 6	19:02	faillog
rw-r--r--	1 root	root	3.8K	Oct 19	22:33	fontconfig.log
rw-r--r--	2 root	root	4.0K	Aug 5	2015	fsck
rw-r--r--	2 root	root	4.0K	Apr 21	2016	gdm3
rw-r--r--	1 root	root	2.4K	Dec 23	22:08	gpu-manager.log
rw-r--r--	3 root	root	4.0K	Aug 5	2015	hp
rw-r--r--	2 root	root	4.0K	Dec 24	02:29	installer
rw-r--r--	1 syslog	adm	21G	Dec 24	02:47	kern.log
rw-r--r--	1 root	utmp	286K	Dec 24	02:47	kern.log
rw-r--r--	2 root	root	4.0K	Dec 24	02:47	kern.log
rw-r--r--	2 mysql	adm	4.0K	Dec 24	02:47	kern.log
rw-r--r--	1 mysql	adm	4.0K	Dec 24	02:47	kern.log
rw-r--r--	1 mysql	adm	4.0K	Dec 24	02:47	kern.log
rw-r--r--	1 root	root	9.0K	Dec 24	02:47	kern.log
rw-r--r--	1 root	root	9.0K	Dec 24	02:47	kern.log
rw-r--r--	1 root	root	9.0K	Dec 24	02:47	kern.log
rw-r--r--	1 root	root	9.0K	Dec 24	02:47	kern.log
rw-r--r--	1 root	root	9.0K	Dec 24	02:47	kern.log
rw-r--r--	2 root	adm	4.0K	Dec 24	02:47	kern.log
rw-r--r--	2 speech-dispatcher	root	4.0K	Dec 24	02:47	kern.log
rw-r--r--	1 syslog	adm	5.0K	Dec 24	02:47	kern.log
rw-r--r--	1 syslog	adm	1.0K	Dec 24	02:47	kern.log
rw-r--r--	2 root	adm	4.0K	Dec 24	02:47	kern.log
rw-r--r--	2 root	adm	4.0K	Dec 24	02:47	kern.log
rw-r--r--	2 root	root	1.0K	Dec 24	02:47	kern.log
rw-r--r--	1 root	utmp	14.0K	Dec 24	02:47	kern.log
rw-r--r--	1 root	utmp	16.0K	Dec 24	02:47	kern.log
rw-r--r--	1 root	root	3.0K	Dec 24	02:47	kern.log
rw-r--r--	1 root	root	31K	Dec 23	20:36	xorg.0.log.0ld
rw-r--r--	1 root	root	31K	Nov 11	22:30	xorg.1.log

Astuce

N'attendez pas la fin de la présentation pour en révéler l'objet.

Parlez de votre produit ou concept (ici, une application de traduction) dès le départ.



2. A la demande

Si vous voulez suivre l'activité à un temps T
d'un système ou d'un service

- **Logiciels d'affichage en direct "système"**
Ils affichent des informations sur un ou des éléments de manière globale
- **Console ou page d'état d'un service**
Affichent différentes infos sur l'état actuel (parfois un peu plus)

Barre de titre: Gestionnaire des tâches

Menu: Affichage

Onglets: Performance | Historique des applications | Démarrage | Utilisateurs | Détails | Services

Statut: 40% 73% 0% 0%

Processus (6)

Processus	Statut	Processeur	Mémoire	Disque	Réseau	Co
Explorateur du Registre		0%	0,9 Mo	0 Mo/s	0 Mbits/s	Très
Explorateur Windows (3)		0,6%	27,3 Mo	0 Mo/s	0 Mbits/s	Très
Gestionnaire des tâches		12,8%	18,1 Mo	0 Mo/s	0 Mbits/s	Faible
Explorateur des applications HTML de ...		0%	1,3 Mo	0 Mo/s	0 Mbits/s	Très
Explorateur de fichiers		14,7%	34,8 Mo	0 Mo/s	0 Mbits/s	Faible
Explorateur de fichiers		0%	0,6 Mo	0 Mo/s	0 Mbits/s	Très
Explorateur de fichiers		0%	5,1 Mo	0 Mo/s	0 Mbits/s	Très
Explorateur de fichiers		0%	88,2 Mo	0 Mo/s	0 Mbits/s	Très
Explorateur de fichiers		0%	5,6 Mo	0 Mo/s	0 Mbits/s	Très
Explorateur de fichiers		0%	1,1 Mo	0 Mo/s	0 Mbits/s	Très
Explorateur de fichiers		0%	0,8 Mo	0 Mo/s	0 Mbits/s	Très

Attention

L'affichage détails est le seul à afficher tous les processus..

Gestionnaire des tâches

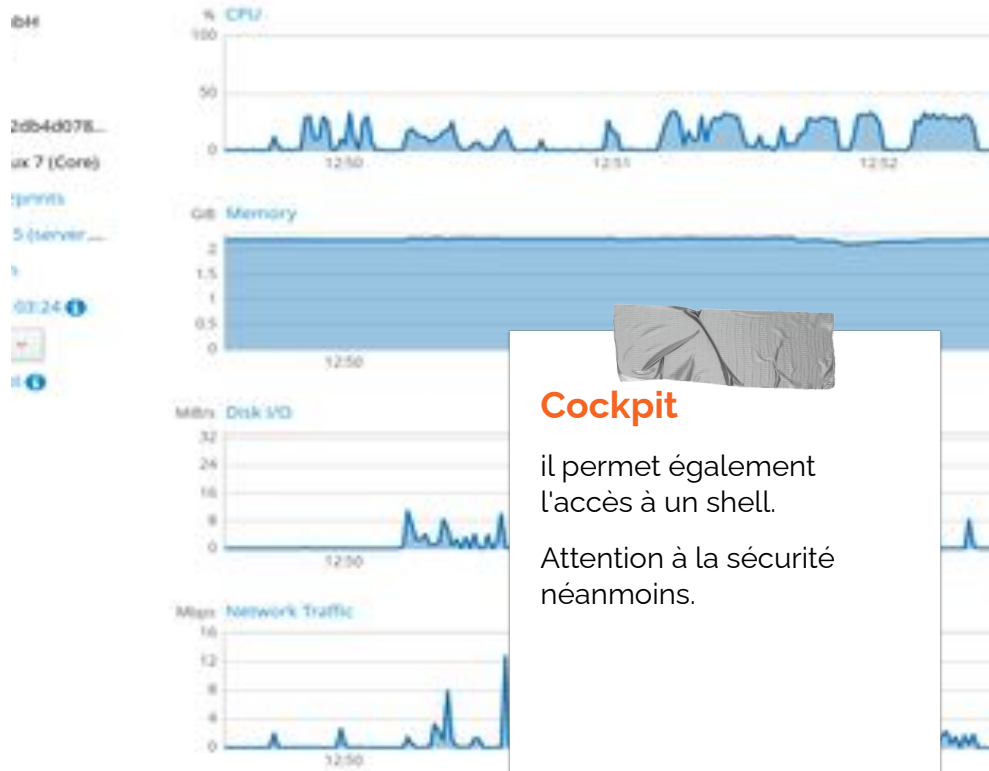
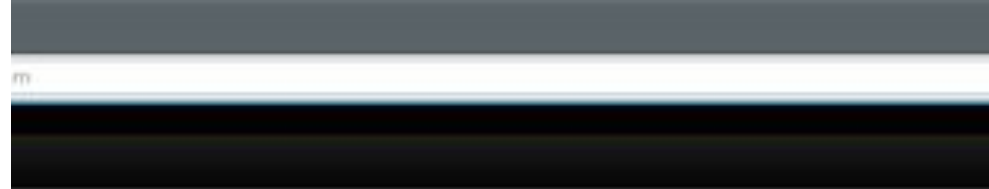
Donne des infos et permet de trier par rapport aux colonnes affichées.

(Windows)

Top / Cockpit

En ligne de commande "top" affiche différents paramètres


Cockpit donne la même possibilité (entre autres), via l'interface web



Cockpit

il permet également l'accès à un shell.

Attention à la sécurité néanmoins.



Il faut lier les
différentes sources
pour arriver à faire le
chemin global.

**Trop
d'informations...**

Tue l'information !

Les expressions régulières

- Définition d'un filtre
- de simple à très complexe
- imbriquable



La criticité/sévérité

DEBUG : information très détaillée pour un but de débogage

INFO : information de suivi

WARN : Suspicion d'événement non désiré

ERROR : Erreur entraînant une rupture de service

Le triage

- Date
- Source (Utilisateur?)
- Sévérité
- Processus / Programme



L'arme principale de la résolution de problèmes ou d'incident

Sans ces fichiers point de salut !

La reproduction d'un problème n'est pas toujours facile.



Quel logs ?

Commencez par les logs de haut niveau pour descendre ou l'inverse. Choisissez un sens et n'oubliez pas le facteur physique (câble débranché)



3. Analyse

La récurrence, la répétition, les liens entre les événements sont les bases de la recherche de preuves et d'explication de comportement curieux.

Il existe des outils qui se basent sur ces fichiers pour en :

- **Tirer des informations**
Exemple : Analyse des statistiques web
- **Informers**
Envoyer une notification dans des cas précis
- **Protéger / Corriger**
Exécuter des actions correctives ou préventives dans certains cas

—

Il y a pléthores de fichiers journaux et d'informations sur l'état d'un service ou d'un machine

L'utilisation de systèmes de centralisation est primordiale



Dans le futur

L'IA a un rôle à jouer : des outils déjouant les attaques web en amont...

Le nombre de fichiers à analyser
explose avec le nombre de service,
l'unification dans des service
d'analyse spécifique est
incontournable pour certaines
entreprises

C'EST LE NERF DE LA GUERRE POUR VALORISER LES DONNÉES.



IA IA IA

Un grand nombre de
startup se lance dans
l'exploitation et
l'interprétation de fichier
journaux.



4. Conclusion

Vous avez maintenant un aperçu

- **De couches différentes**
De la connection à l'utilisation d'application web
- **De sources différentes**
- **Qui sont liées entres elles**
La connexion physique donne accès à un service qui utilise un autre service...

Analyse

Tâches de fond

Des déclencheurs, actions,
etc

Liens entre sources

par date / utilisateur

Préventive

Corrective

Fichiers

exploitations du contenu

Modification

Ajout de contenu pour
surveillance / vérification
après correction

Sans les journaux

Combien de personnes ont accédé à notre site web ?

euh....

Qui s'est connecté hier soir sur l'admin ?

Je sais pas...

Pourquoi le site n'est plus accessible, et depuis combien de temps ?

Attends je vais voir...

Bref, c'est chaud !

Expressions régulières

Structure

Comment se décrit un modèle

cool : recherche les lignes avec cool

CooL : les lignes avec CooL

Co*L : les lignes avec C puis de 0 à plusieurs o puis L

Co+L : les lignes avec C puis de 1 à plusieurs o puis L

(CooL){4,5} : les lignes avec de 4 à 5 fois **CooL**

et le reste ?

[Regular Expressions Cheat Sheet by DaveChild - Download free from Cheatography](#)

Anchors

<code>^</code>	Start of string, or start of line in multi-line pattern
<code>\A</code>	Start of string
<code>\$</code>	End of string, or end of line in multi-line pattern
<code>\Z</code>	End of string
<code>\b</code>	Word boundary
<code>\B</code>	Not word boundary
<code>\<</code>	Start of word
<code>\></code>	End of word

Character Classes

<code>\c</code>	Control character
<code>\s</code>	White space
<code>\S</code>	Not white space
<code>\d</code>	Digit
<code>\D</code>	Not digit

Quantifiers

<code>*</code>	0 or more	<code>{3}</code>	Exactly 3
<code>+</code>	1 or more	<code>{3,}</code>	3 or more
<code>?</code>	0 or 1	<code>{3,5}</code>	3, 4 or 5

Add a `?` to a quantifier to make it ungreedy.

Escape Sequences

<code>\</code>	Escape following character
<code>\Q</code>	Begin literal sequence
<code>\E</code>	End literal sequence

"Escaping" is a way of treating characters which have a special meaning in regular expressions literally, rather than as special characters.

Common Metacharacters

<code>^</code>	<code>[</code>	<code>.</code>	<code>\$</code>
<code>{</code>	<code>*</code>	<code>(</code>	<code>\</code>

Groups and Ranges

<code>.</code>	Any character except new line (<code>\n</code>)
<code>(a b)</code>	a or b
<code>(...)</code>	Group
<code>(?:...)</code>	Passive (non-capturing) group
<code>[abc]</code>	Range (a or b or c)
<code>[^abc]</code>	Not (a or b or c)
<code>[a-q]</code>	Lower case letter from a to q
<code>[A-Q]</code>	Upper case letter from A to Q
<code>[0-7]</code>	Digit from 0 to 7
<code>\x</code>	Group/subpattern number "x"

Ranges are inclusive.

Pattern Modifiers

<code>g</code>	Global match
<code>i *</code>	Case-insensitive
<code>m *</code>	Multiple lines



Et maintenant un peu de réflexion avec de la pratique.



**TD/TP Expression
régulière**

C'est parti : [Bloc 1 - TP -
Expression régulière](#)

One more time.



TD/TP

[Bloc 1 - TP -
Manipulation de fichier
sous linux](#)