I] Installation de Splunk Enterprise sous Windows (usage perso)

0- on n'est pas obligé d'avoir une bête de course pour installer Splunk Enterprise pour découverte / démo en utilisation mono-utilisateur :

https://docs.splunk.com/Documentation/Splunk/8.1.0/Capacity/Referencehardware

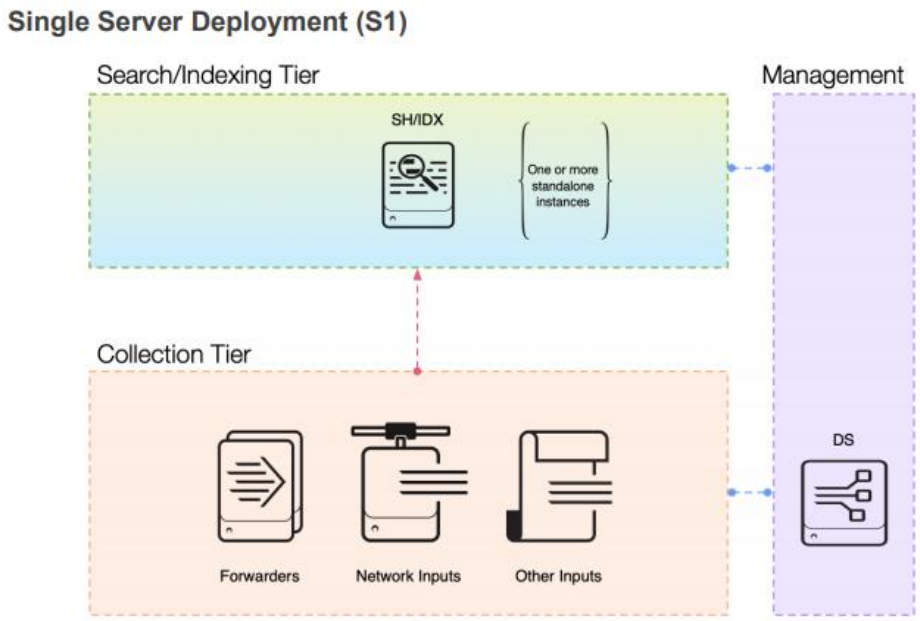
Malgré tout :

Reference host specification for single-instance deployments

This represents the minimum basic instance specifications for a production grade Splunk Enterprise deployment. A single-instance represents an S1 architecture in SVA:

- An x86 64-bit chip architecture
- 12 physical CPU cores, or 24 vCPU at 2Ghz or greater speed per core.
- 12GB RAM.
- For storage, see What storage type should I use for a role?
- A 1Gb Ethernet NIC, optional 2nd NIC for a management network .
- A 64-bit Linux or Windows distribution. See Supported Operating Systems in the *Installation Manual*.

SVA = Splunk Validated Architecture : https://www.splunk.com/pdfs/technical-briefs/splunk-validated-architectures.pdf



Spécifications de l'appareil

Nom de l'appareil	DESKTOP-N9FH26V
Processeur	Intel(R) Core(TM) i7-4800MQ CPU @ 2.70GHz 2.70 GHz
Mémoire RAM installée	32,0 Go
ID de périphérique	D84EC0CC-153B-4CD2-B382-58835B102CA5
ID de produit	00331-60000-00000-AA870
Type du système	Système d'exploitation 64 bits, processeur x64
Stylet et fonction tactile	La fonctionnalité d'entrée tactile ou avec un stylet n'est pas disponible sur cet écran

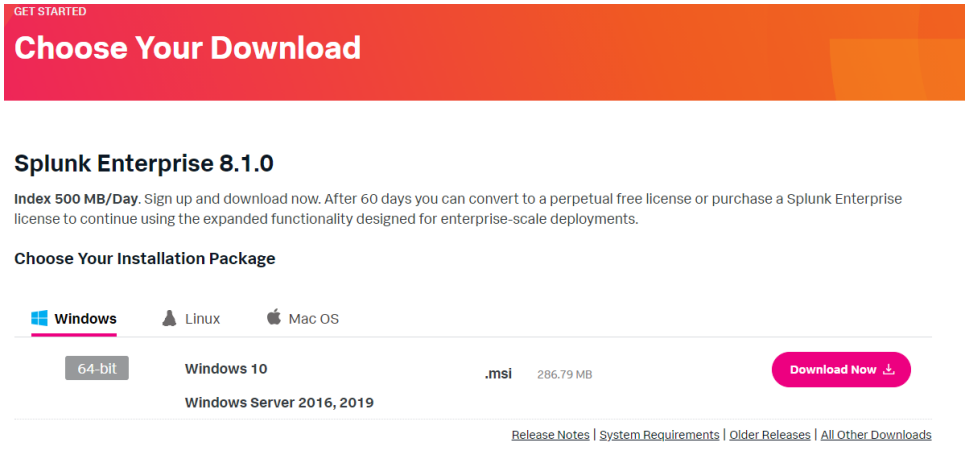
Renommer ce PC

Spécifications de Windows

Édition	Windows 10 Professionnel N
Version	1909
Installé le	15/02/2020
Version du système d'exploitation	18363.1198

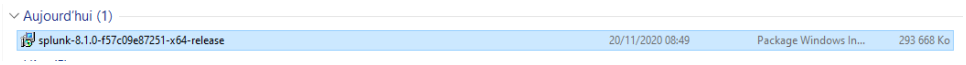
1- on télécharge le msi sur le site SPLUNK (demande création de compte – parfois pas possible avec mails exotiques – Peu de spams)

https://www.splunk.com/en_us/download/splunk-enterprise.html?utm_campaign=bing_emea_tier1_en_search_brand&utm_source=bing&utm_medium=cpc&utm_term=%2Bsplunk%20%2Bdownload&utm_content=Splunk_Enterprise_Demo&_bt=71193491264909

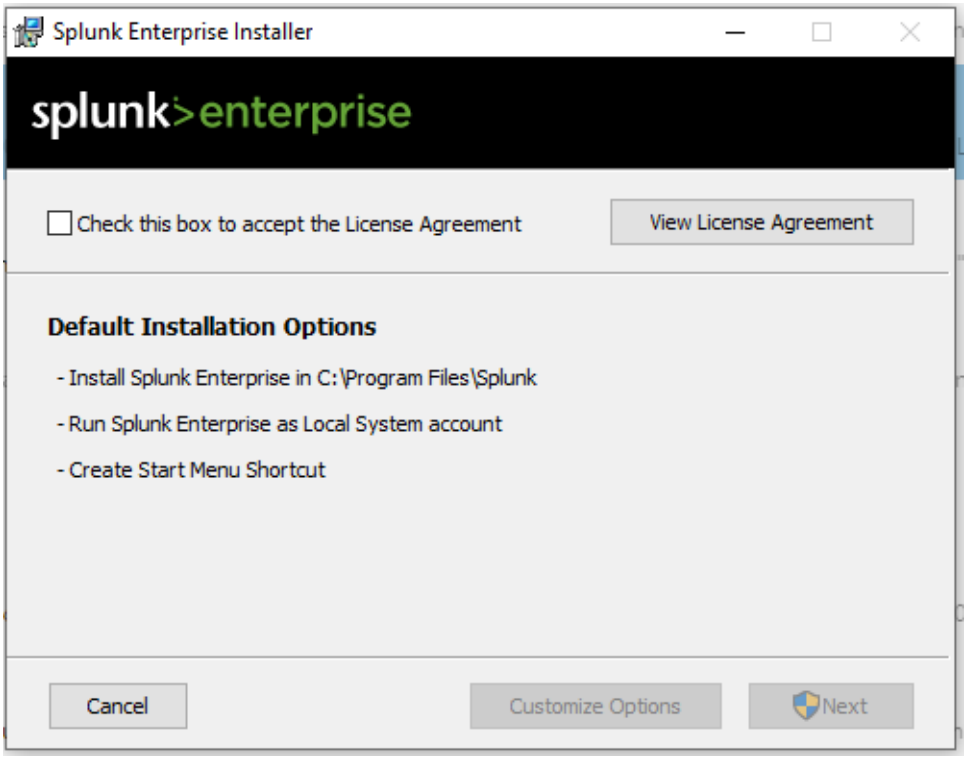


Voir pour les requirements : https://docs.splunk.com/Documentation/Splunk/latest/Installation/Systemrequirements#Recommended_hardware

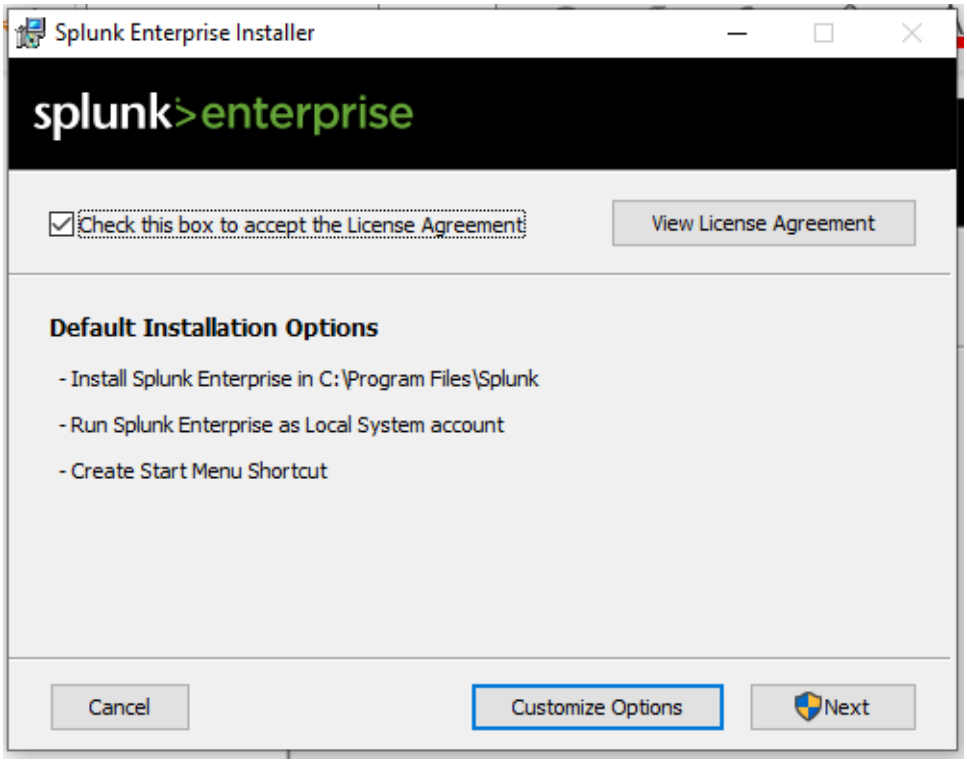
Et possibilité d'installer VMs (attention baisse de perf) ou images Docker (https://hub.docker.com/r/splunk/splunk/.)



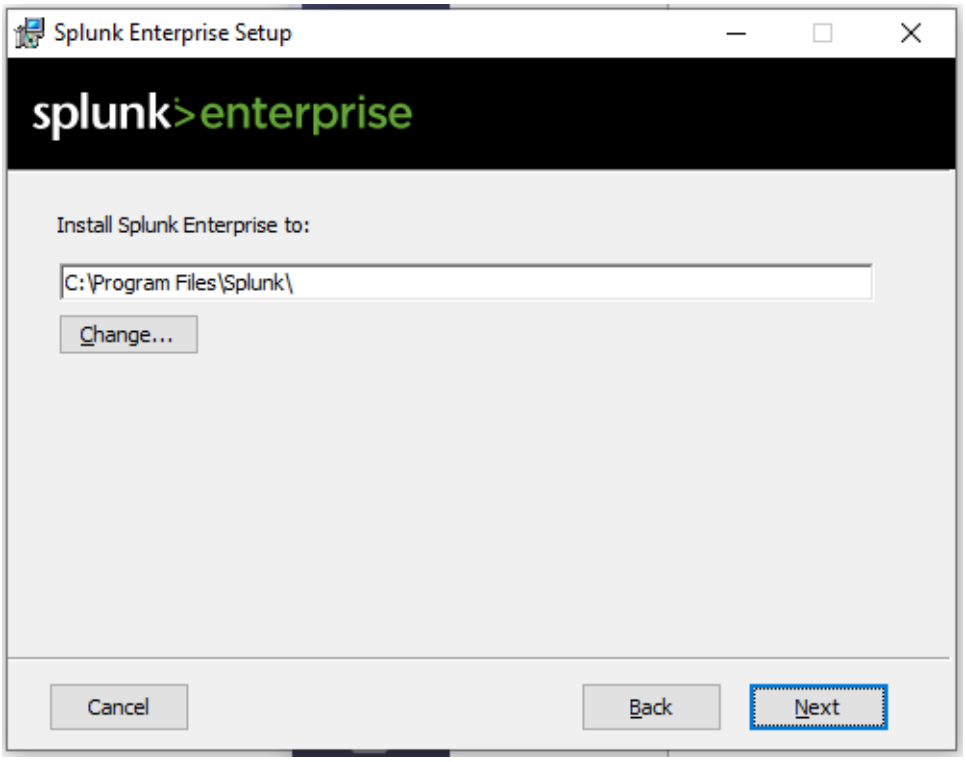
Traditionnellement, on fait clic, clic, clic et ça a l'air de marcher … mais on n'a rien compris.



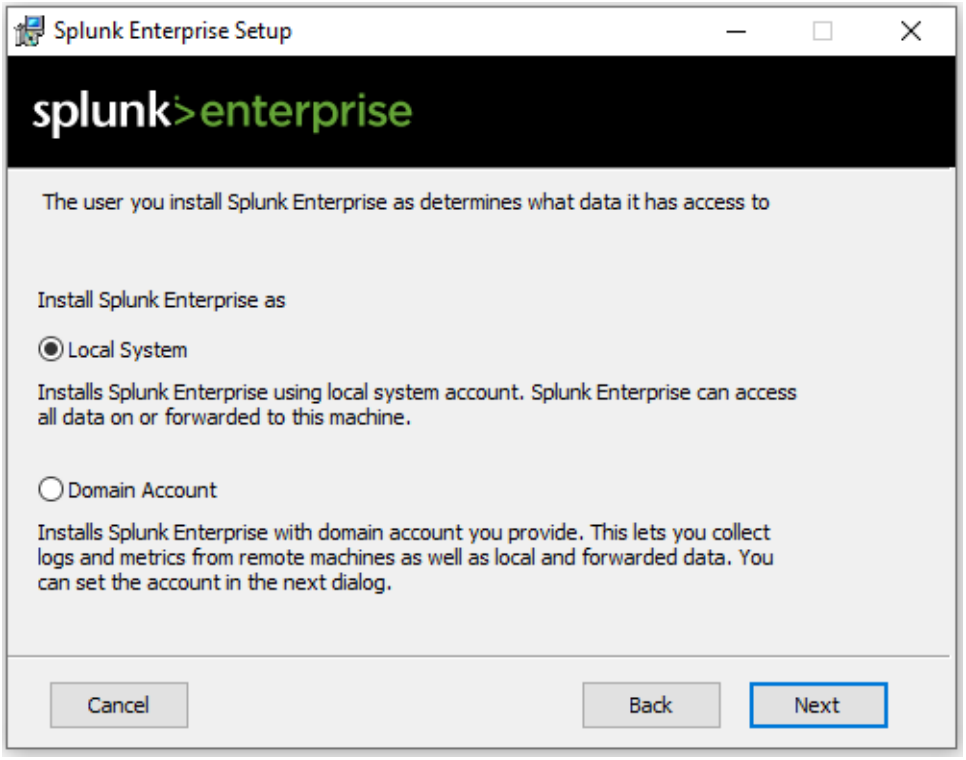
Je ne sais pas à quoi on s'engage en acceptant la licence, mais ça ne fait pas de mal de la transmettre au service juridique de votre structure. Quoi qu'il en soit, l'acceptation est obligatoire.



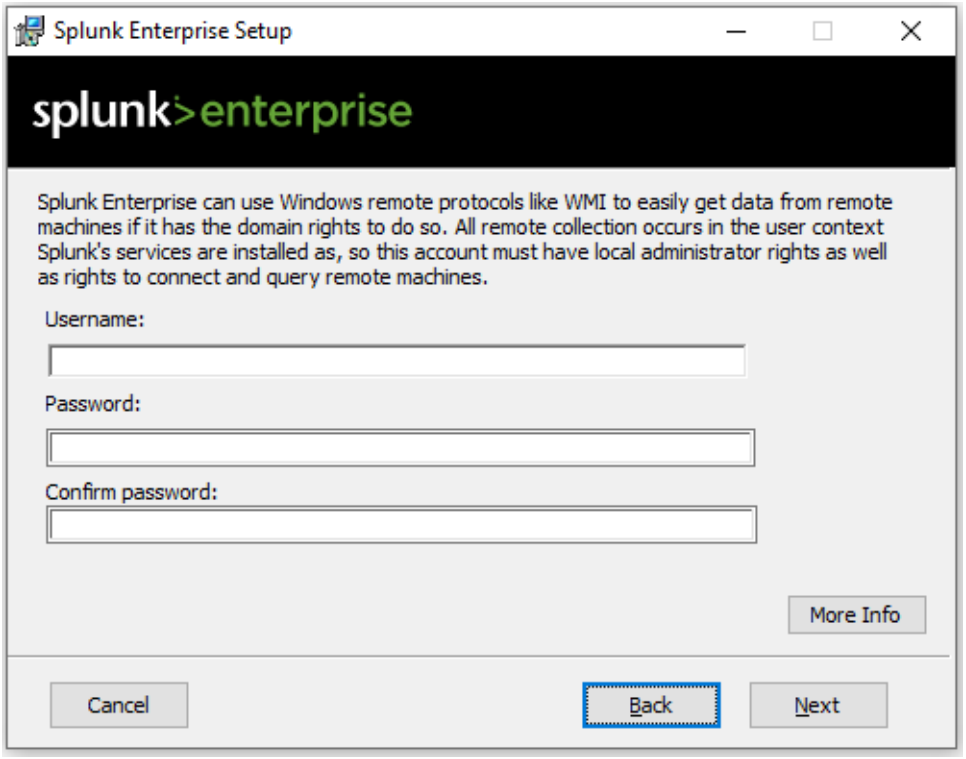
Il est indispensable de passer par “Customize Options” pour un minimum de maîtrise.

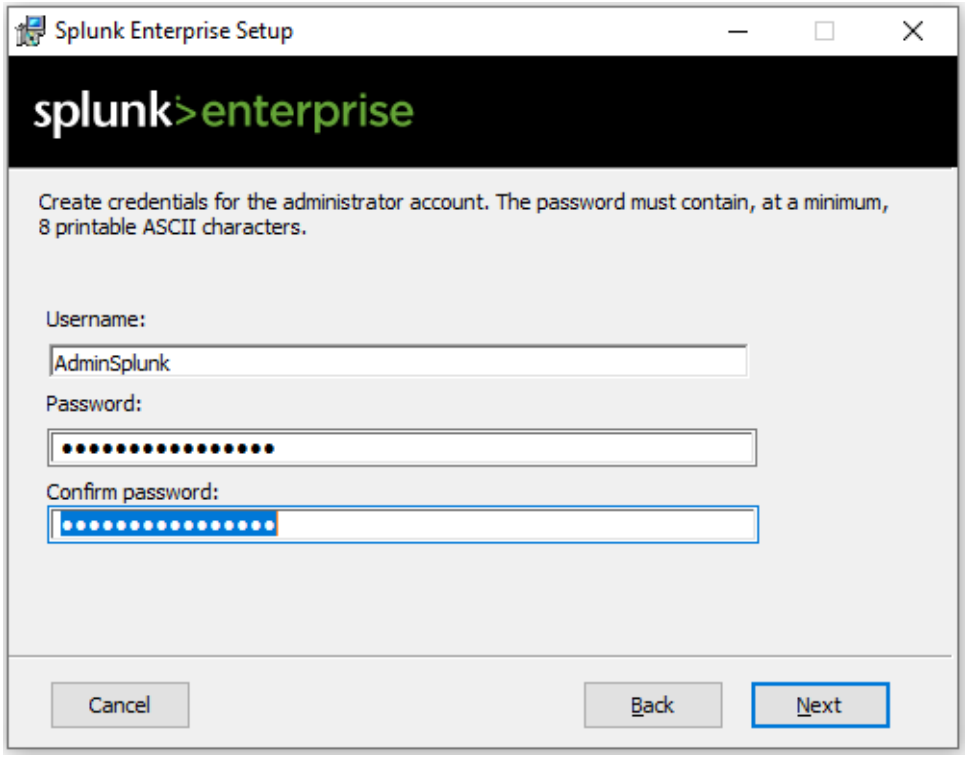


Chemin d’installation selon vos contraintes personnelles ou d’entreprise.

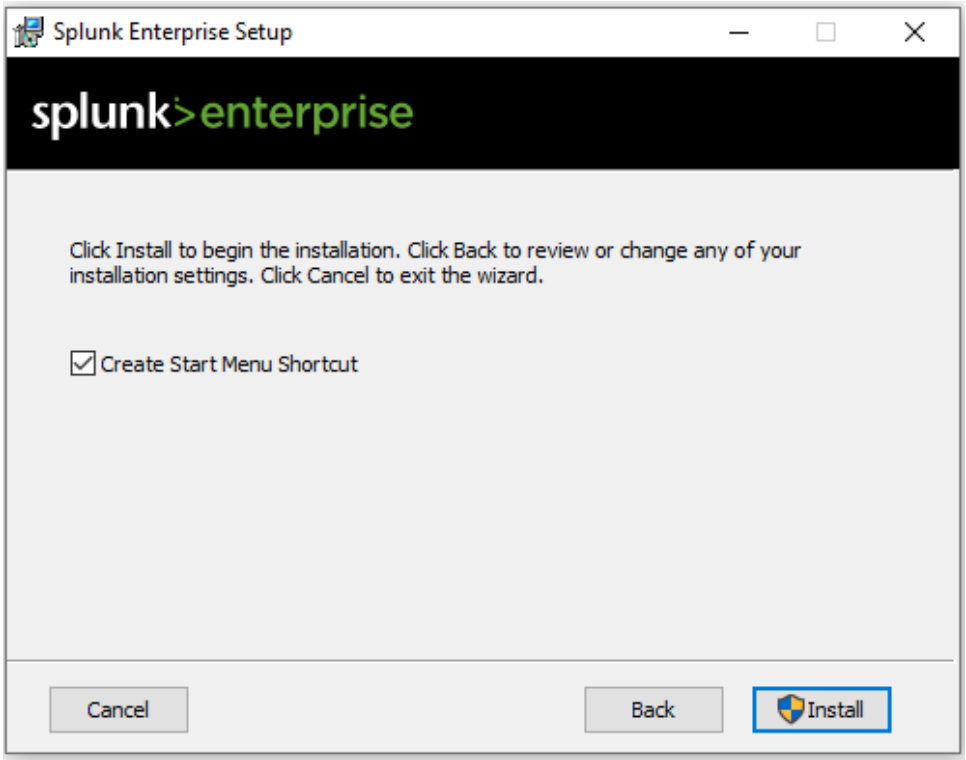


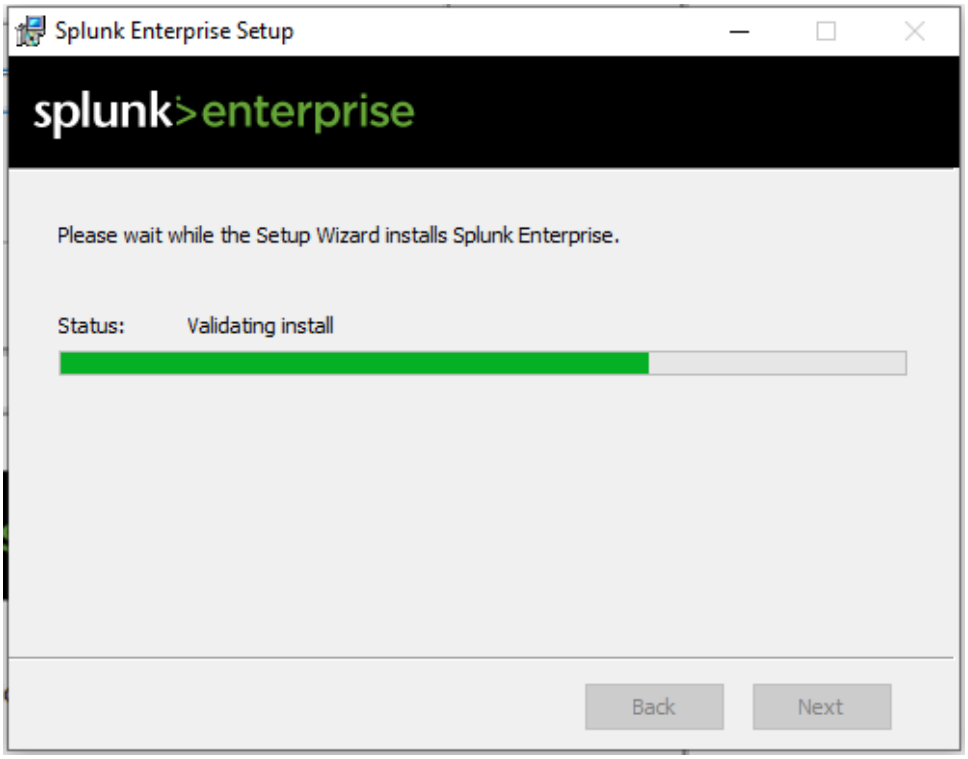
Choix du compte d'installation (souvent local … car les données des autres serveurs du domaine peuvent être récupérés par d'autres moyens). On se pose la question de déterminer le niveau de droits de ce compte (on en reparlera quand on abordera les Apps)

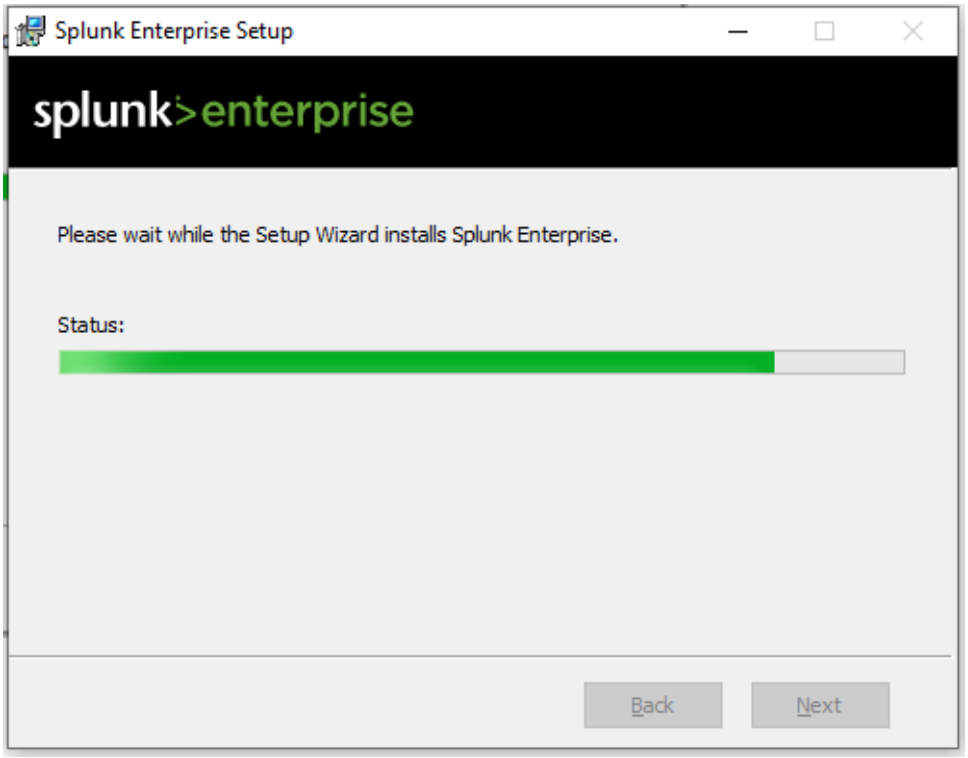


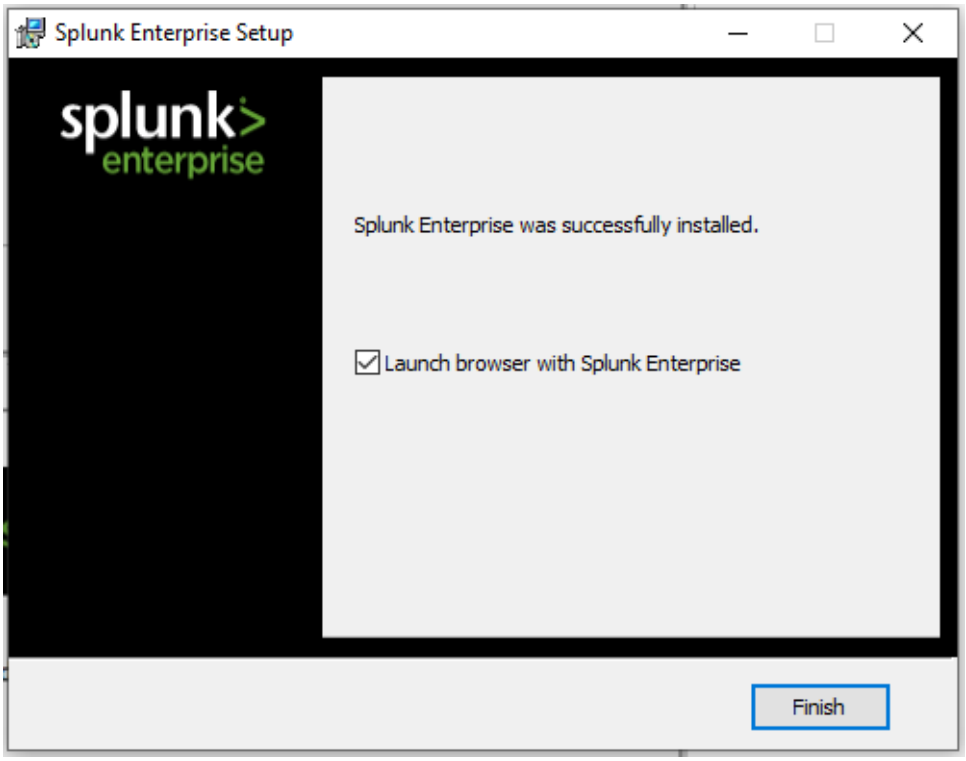


Vos informations de connexion au Keepass avant de les oublier (même s'il existe une procédure un peu tordue pour faire un reset password)

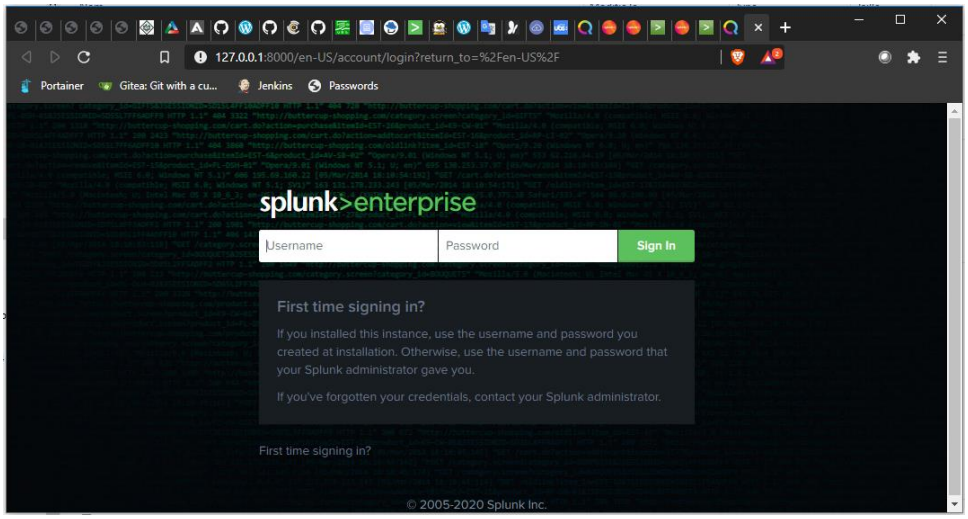



Splunk Enterprise Setup
splunk>enterprise
Please wait while the Setup Wizard installs Splunk Enterprise.
Status: Validating install
Back
Next

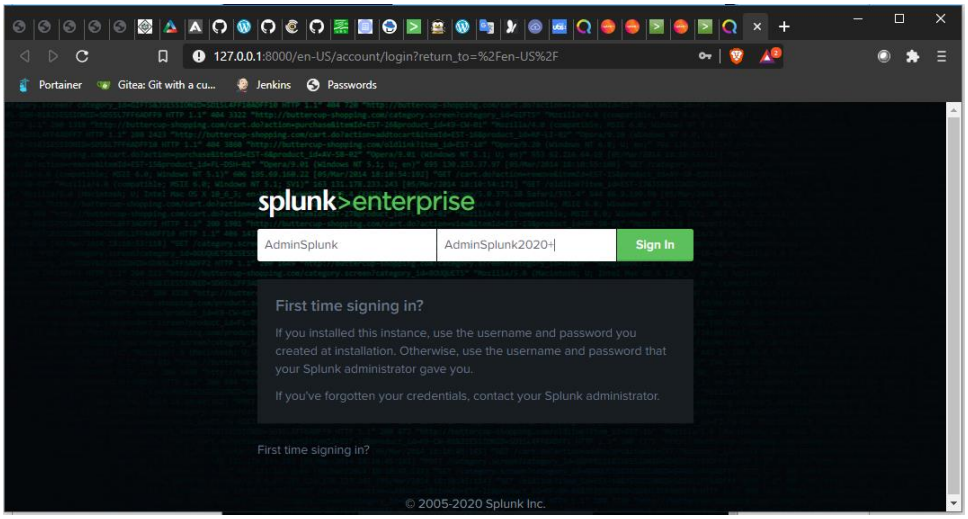

Splunk Enterprise Setup
splunk>enterprise
Please wait while the Setup Wizard installs Splunk Enterprise.
Status:
Back
Next



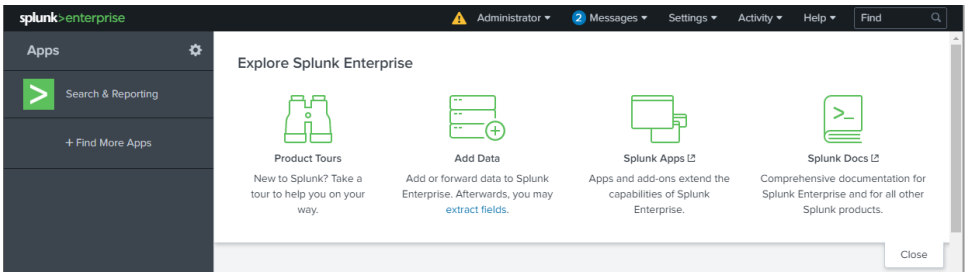
Installation visiblement Ok



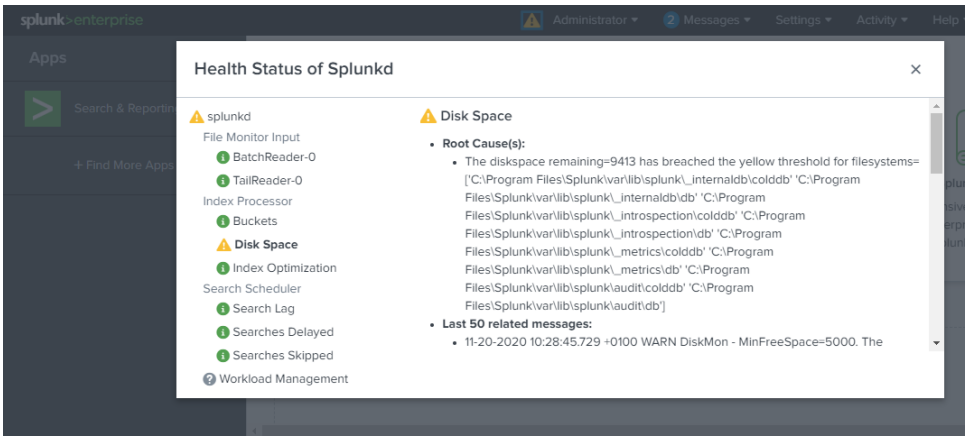
L'interface web s'ouvre automatiquement (localhost, port 8000, http)



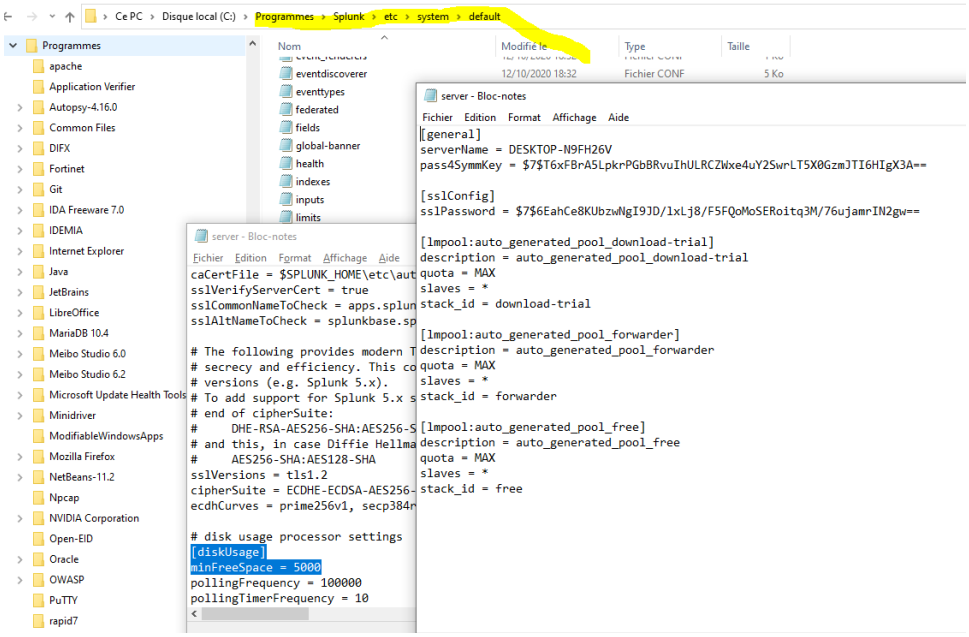
Première connexion



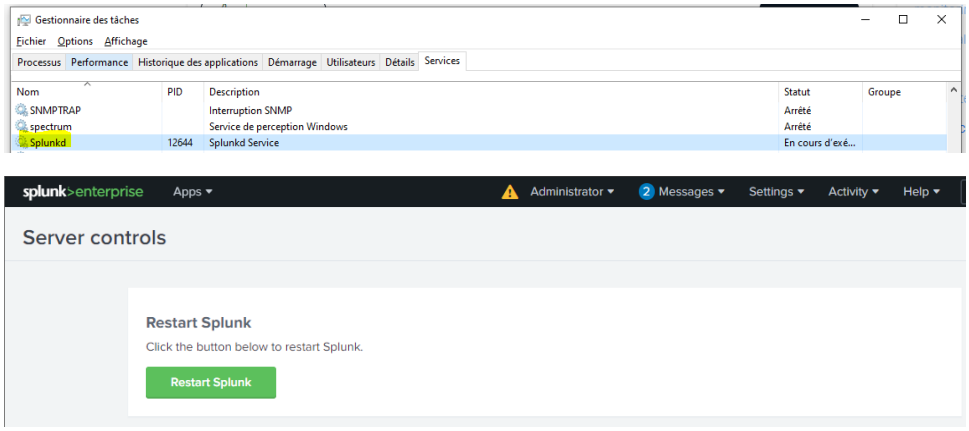
Fonctionnement ok … ou presque (Warning !)



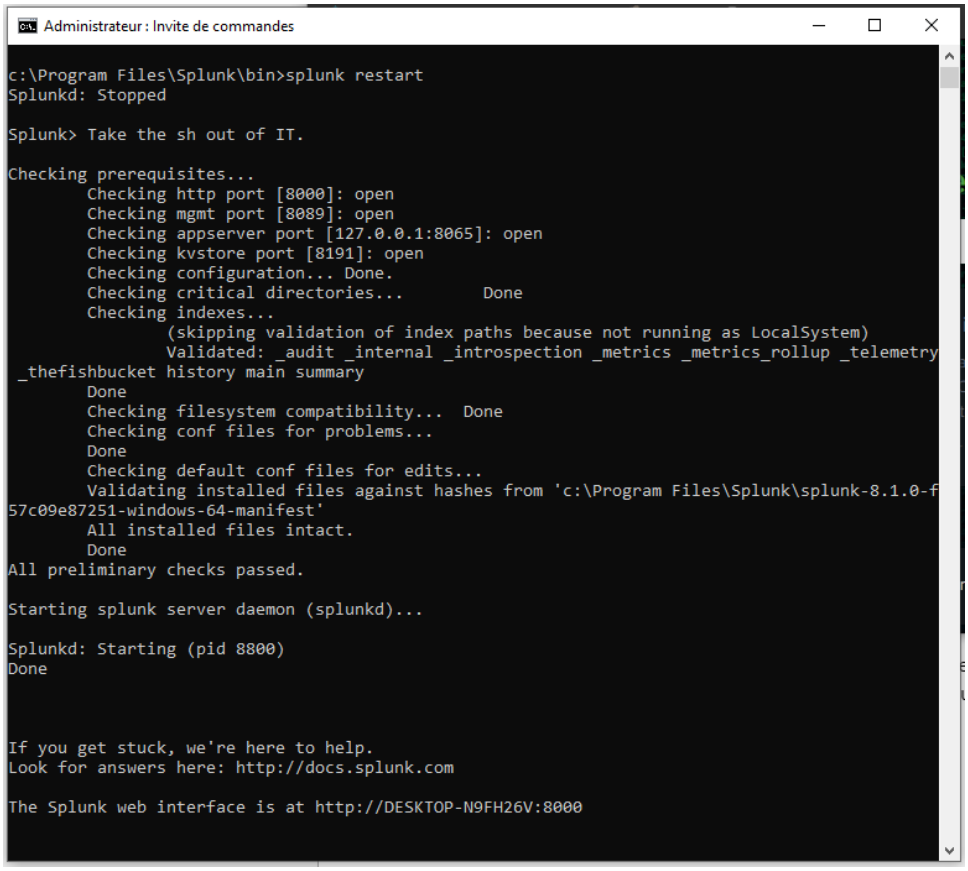
Je dispose de 9 Go de libre sur mon disque soit, moins de 2 fois l’espace disque minimum suffisant.

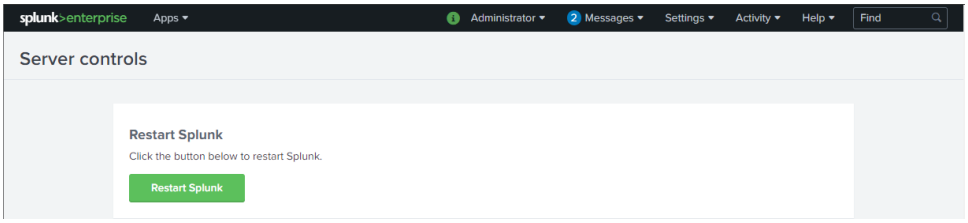


Tout est configurable dans des fichiers default / local !!!!!

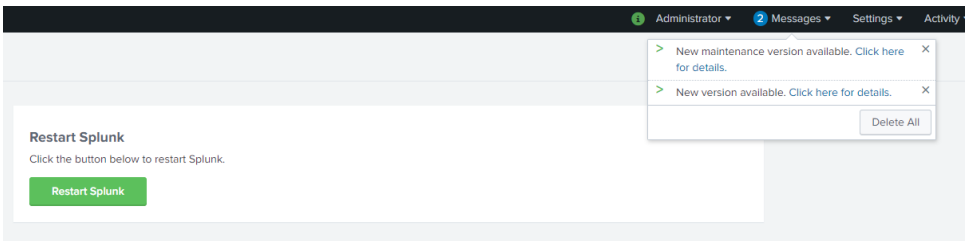


Le redémarrage suite augmentation de l'espace disque s'avère nécessaire. Opération faisable par redémarrage du service ou via GUI ou cmd



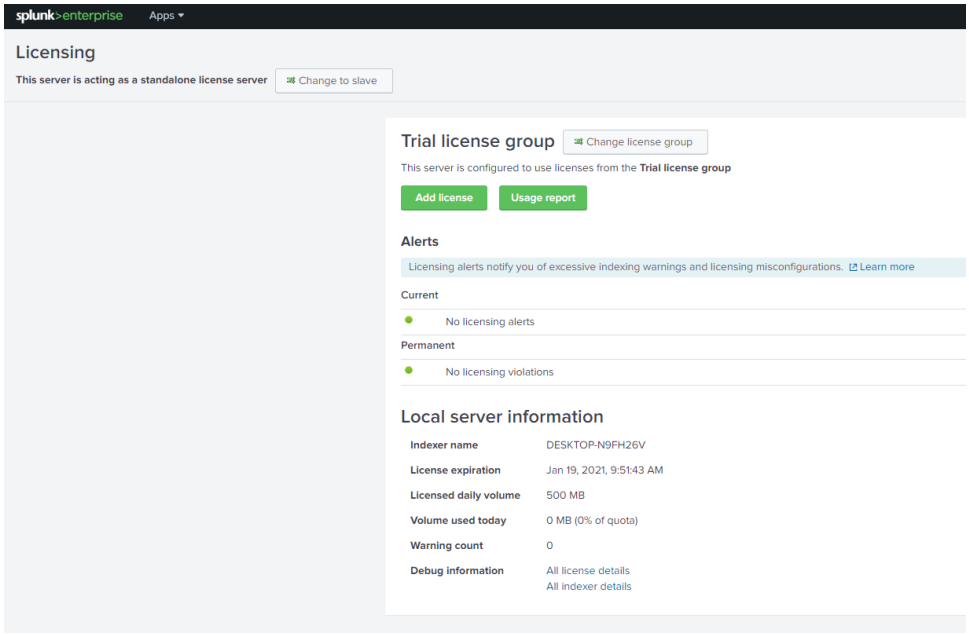


Le warning a disparu. En revanche, nous avons 2 messages

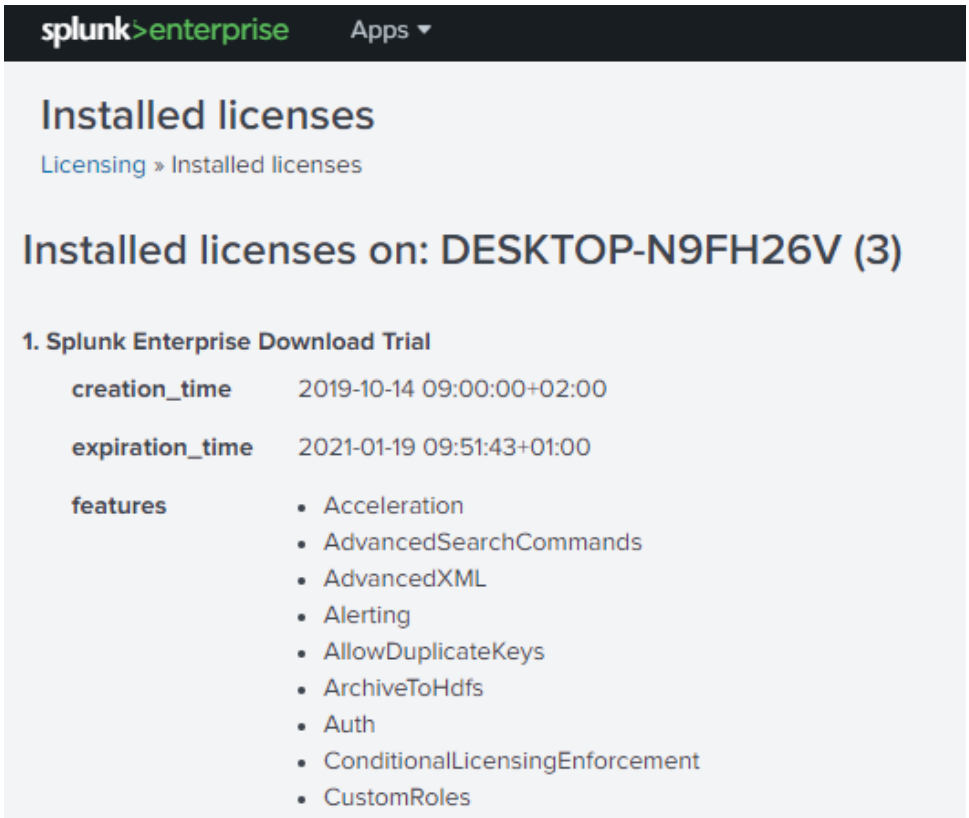


Après vérification, il n'y a pas de nouvelle version :

splunk-8.1.0-f57c09e87251-x64-release

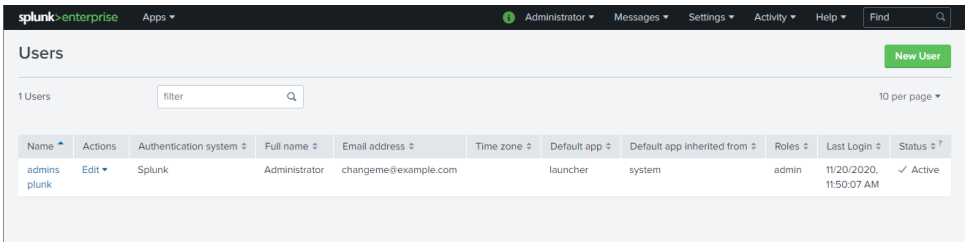


Politique de licence : 500 Mo/j pendant 2 mois. Nous avons droit à 5 dépassements de licence sur 30 jours glissants.

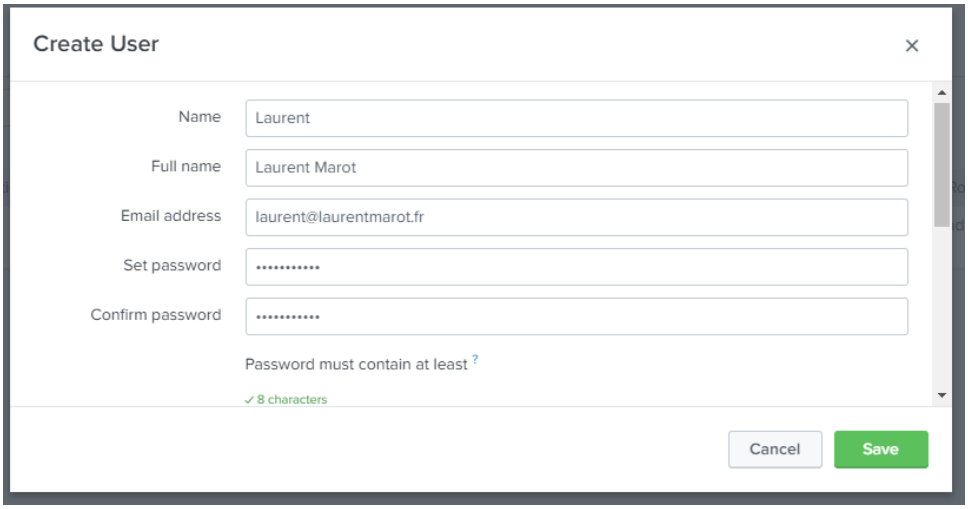


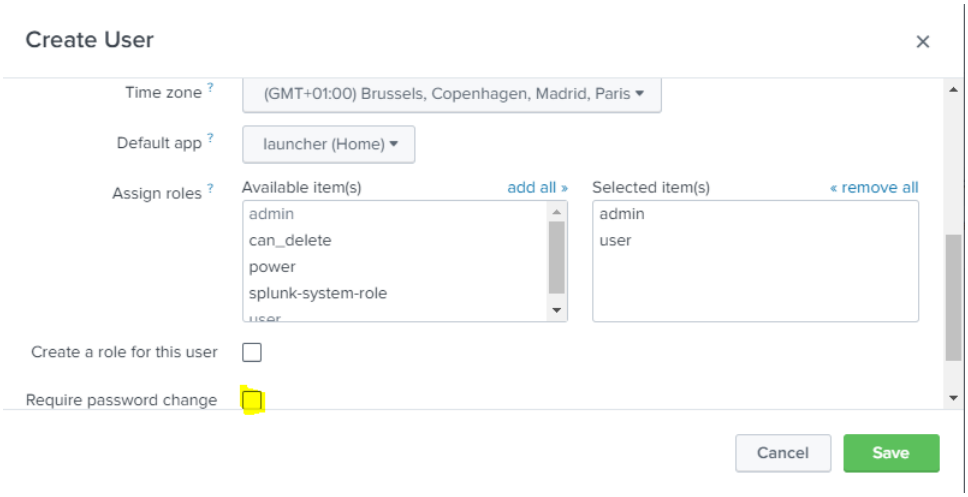
hash	3D2E4634883A0648384BEEDCA226FC78EC2598B547553C90830CA3A93C305886
is_unlimited	False
label	Splunk Enterprise Download Trial
max_violations	5
payload	None
quota_bytes	524288000.0
sourcetypes	
stack_name	download-trial
status	VALID
type	download-trial
window_period	30

En cas de dépassement licence, il vous faudra demander une licence de déblocage au service commercial SPLUNK



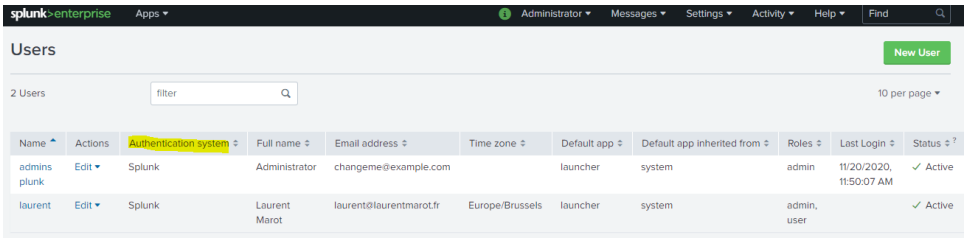
Première règle de sécu, je crée un compte pour un nouvel utilisateur (moi-même) et je mets le compte adminSplunk au coffre. Il est indispensable d'assurer une traçabilité des actions depuis la plateforme.





Je luis donne tous les droits (role admin – nous reviendrons sur cette notion de rôle).

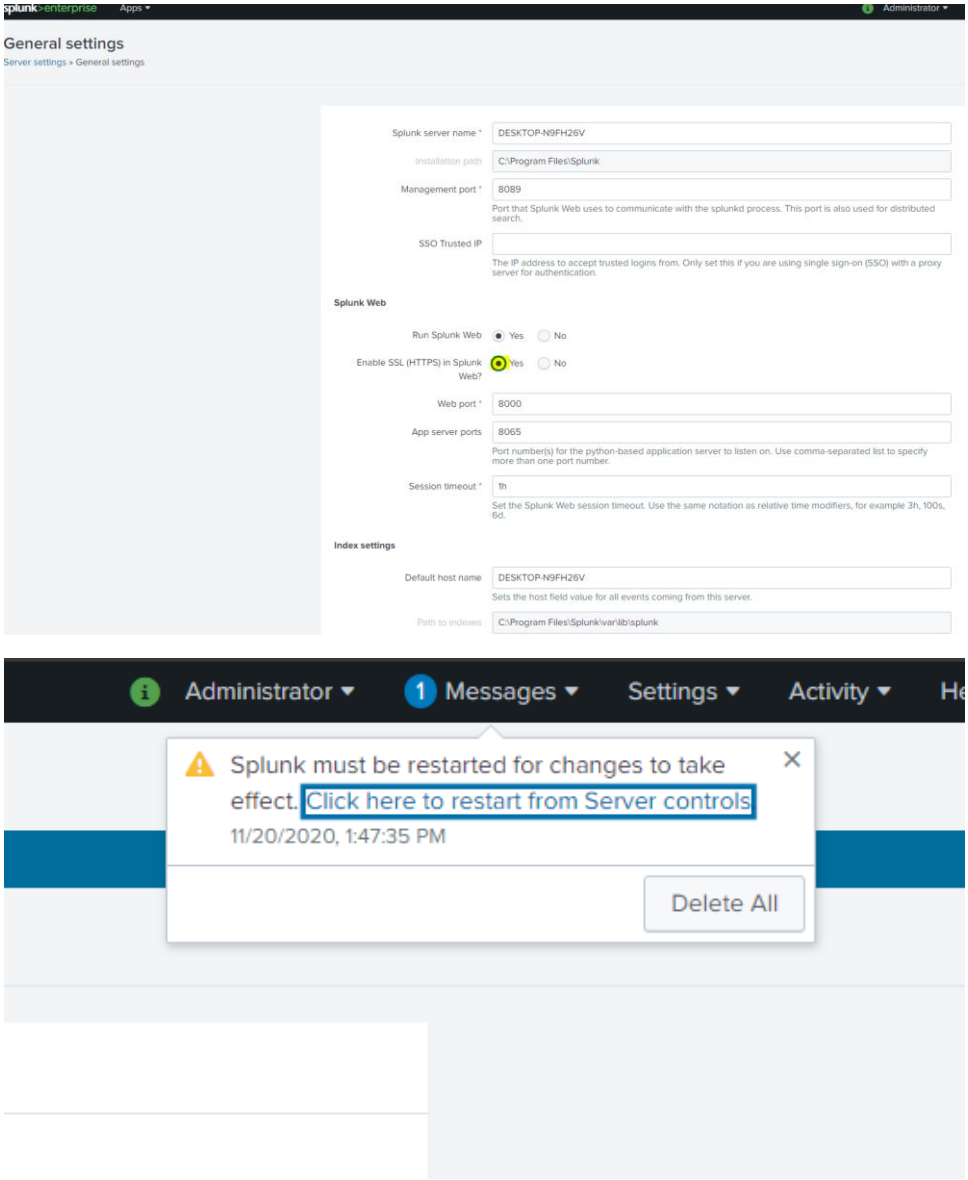
Il va de cela que dans la vraie vie on activerait le “Require password change on first login”



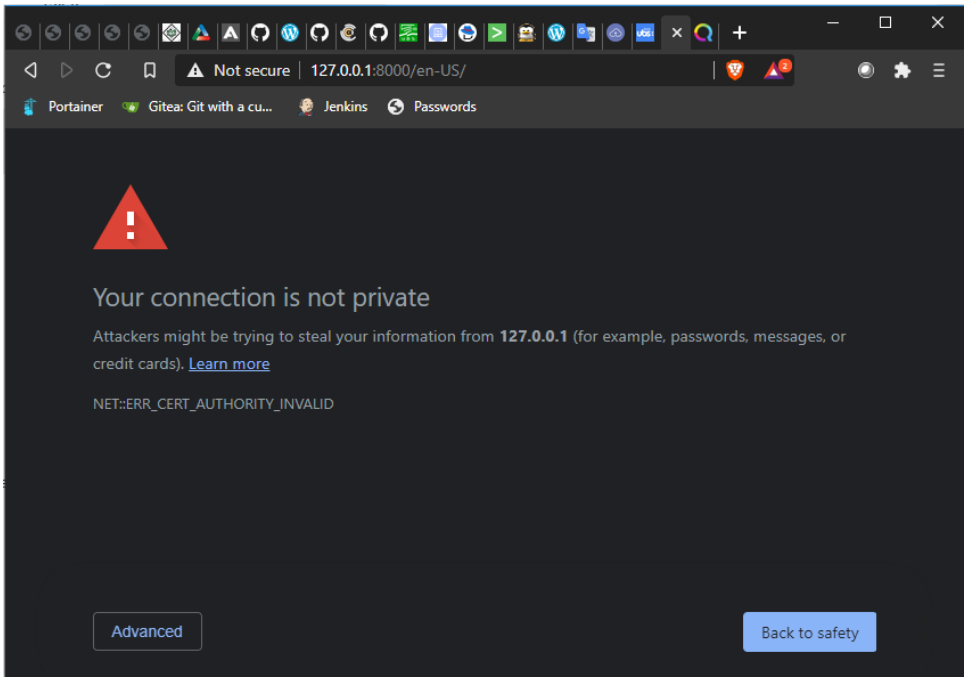
On s’aperçoit qu’il existe plusieurs “Authentication System”. Heureusement, en entreprise cela permet de s’interfacer avec le LDAP.

J’ai renseigné une adresse email mais pas de Gateway SMTP, ce qui devient rapidement indispensable pour tester les notifications d’alertes par messagerie.

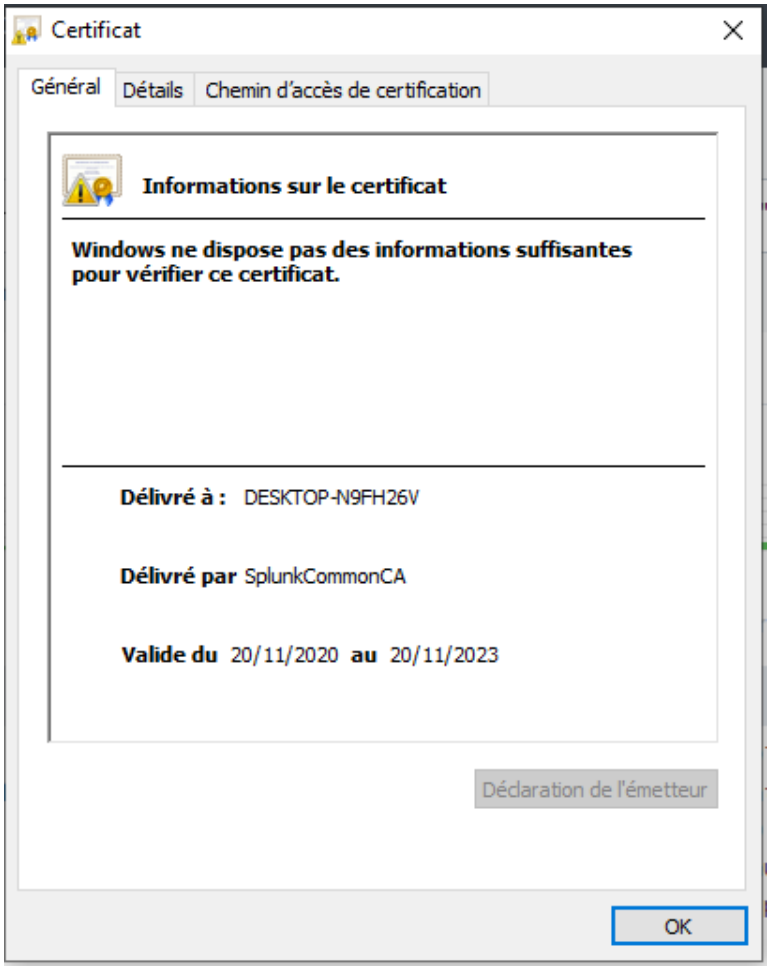
Deuxième action de sécurité : on active HTTPS … au cas où on accède au travers du réseau

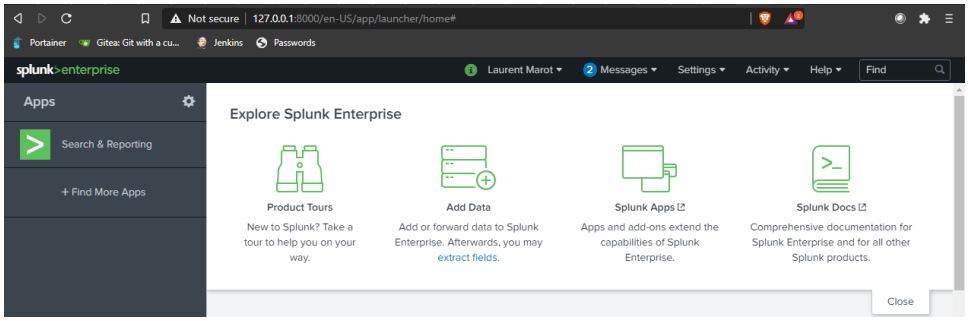


Et on doit alors redémarrer.



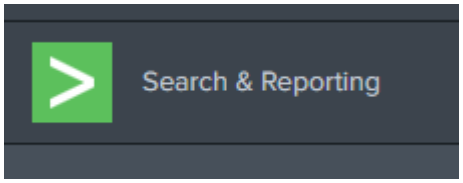
Forcément, certificat signé par CA non reconnue, ça couine un peu … mais au moins le flux est chiffré. Et si on arrive à se connecter avec son compte c'est que c'est certainement la bonne machine.

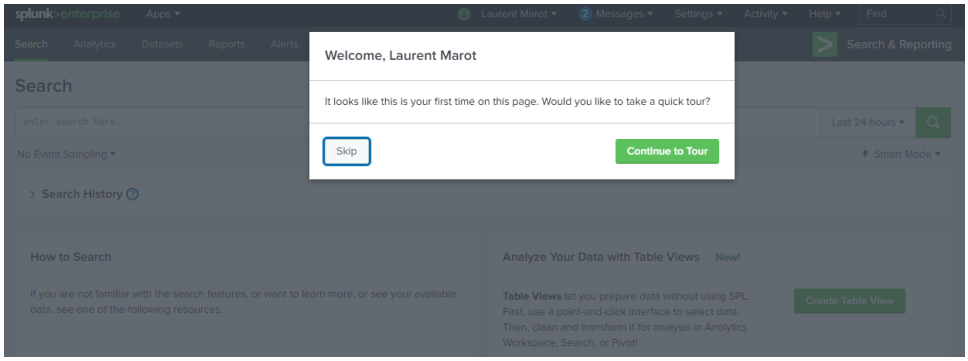




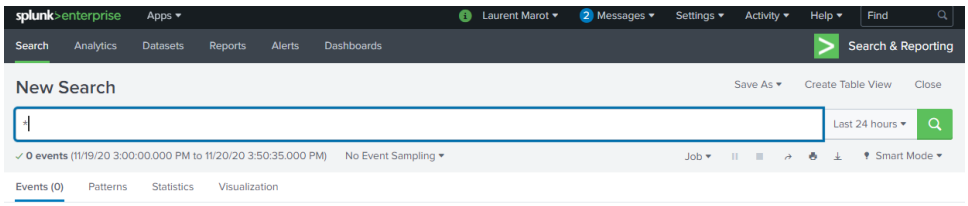
Je suis connecté avec mon nouveau compte.

Voyons ce que la bête a dans le ventre …





N’hésitez pas à parcourir le “Quick Tour”



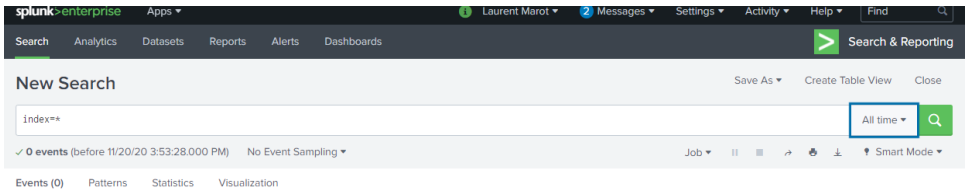
Lançons notre première recherche : *. Magnifique : 0 résultat !

Notons que :

1- on peut passer en langue française

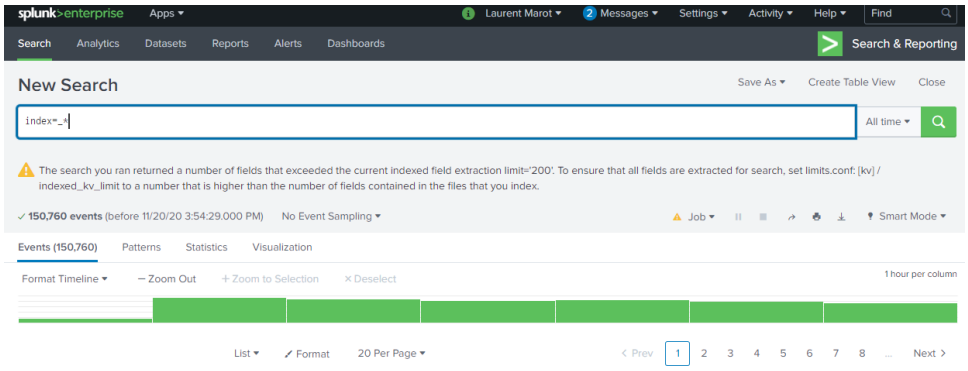
2- la recherche porte sur les 24 dernières heures

3- le mode de recherche est “Smart Mode”

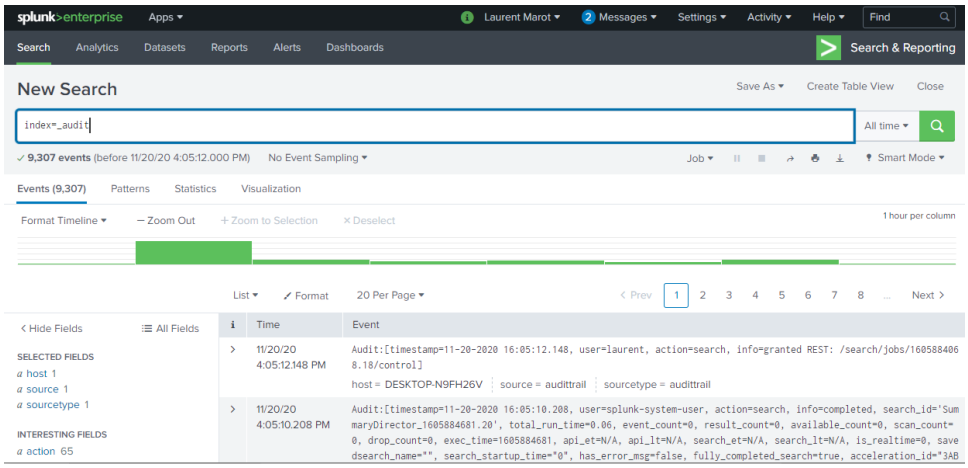


index=* sur “All Times” => pas mieux !

Attention, sur les mots clefs (par exemple “index”) la solution est sensible à la casse.



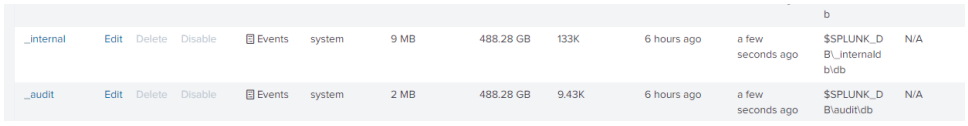
index=_* sur “All Times” => 150 760 enregistrements (et un warning bizarre) !



index=_audit => **9,307 events** (before 11/20/20 4:05:12.000 PM)

Donc, il semblerait que l’on ait des index... par défaut (Vous vous souvenez de la fonction d’indexation ?)

Affirm' … 11 index dont certains sont préfixés par un _



Pour la petite histoire, Splunk, pour des besoins d’audit de son propre fonctionnement indexe ses logs d’activité et d’erreurs dans l’index “_audit”

Et ces index se retrouvent au niveau du système de fichiers. On regardera en détail plus tard.

La gestion de ces index est très importante en termes de performance et en termes de sécurité.

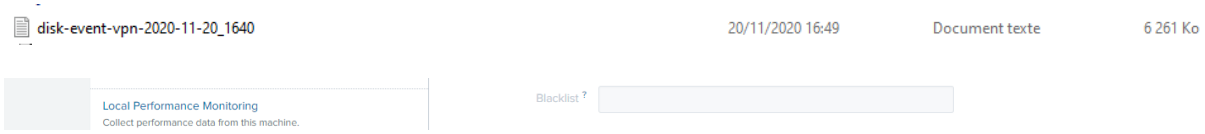
Un exemple de ce qu'il ne faut pas faire : ne pas être vigilant sur l'index qui va accueillir nos données.

Imaginons vouloir alimenter notre outil avec des logs trouvés par hasard …

Add Data

La partie haute de l'écran est nouvelle, je ne sais pas à quoi elle sert. Des exemples j'imagine... il y a de la ressource pléthorique sur Splunk.

Ne nous soucions pas trop des options. Indexons ce fichier :

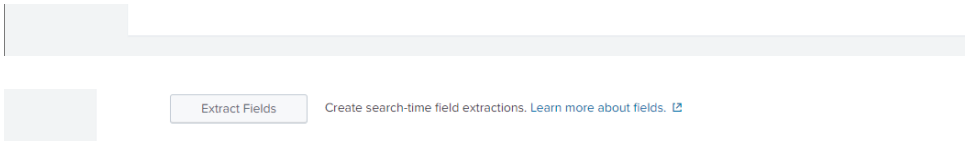


Splunk arrive à trouver la date toute seul (souvenez-vous : un enregistrement <=> 1 bloc de données avec un timestamp. Attention on a souvent du multilignes dans les logs applicatives).

Choisissons, à tort, le source-type à "syslog" dans un premier temps :

Splunk n'apprécie pas le format de date mais ce n'est pas très grave dans notre cas.

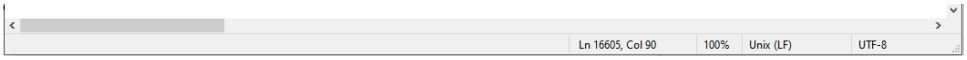
On se dit que l'on jettera un œil à "How do index work ?" quand on aura le temps, et on ne le fera … jamais.



Voilà, on peut ajouter la ligne “Expert Splunk” sur son CV.

On peut lancer notre première recherche.

On nous annonce 16 605 évènements.



Seems to be correct … Notepad me dit la même chose. N’oubliez jamais de contrôler la cohérence de ce que vous annonce l’outil !

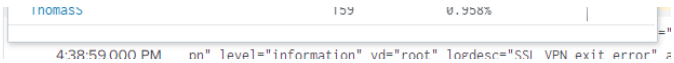
J’ai aussi une cohérence avec ce que m’indiquent les index (si je survole le 16,6k avec la souris, je retrouve bien mes 16 605 enregistrements) … en particulier l’index “main”.

En revanche, lorsque je fais des tests … je n’alimente JAMAIS l’index “main” qui risque autrement de devenir un index poubelle (accessible à tous). J’aurais dû créer un index dédié avec un nom signifiant !

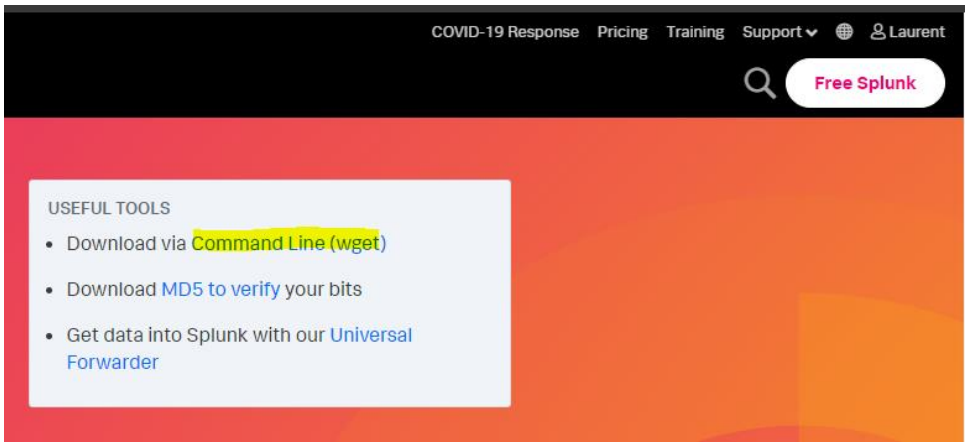
Mais revenons à nos moutons … la touche “back” de votre navigateur est très pratique pour revenir sur l’historique de vos actions.

Le parsing “automatique” de nos logs, malgré un source-type hasardeux, semble s’être bien déroulé.

On va donc pouvoir utiliser les pires aspects de ce genre d’outil pour attribuer nos notes, non plus au hasard cette année mais selon la convention que celui qui se connecte le plus souvent est celui qui travaille le plus, donc est le plus méritant :



II] Installation de Splunk Enterprise Linux



wget -O splunk-8.1.0.1-24fd52428b5a-Linux-x86_64.tgz 'https://www.splunk.com/bin/splunk/DownloadActivityServlet?architecture=x86_64&platform=linux&version=8.1.0.1&product=splunk&filename=splunk-8.1.0.1-24fd52428b5a-Linux-x86_64.tgz&wget=true'

Root ??? Really ???

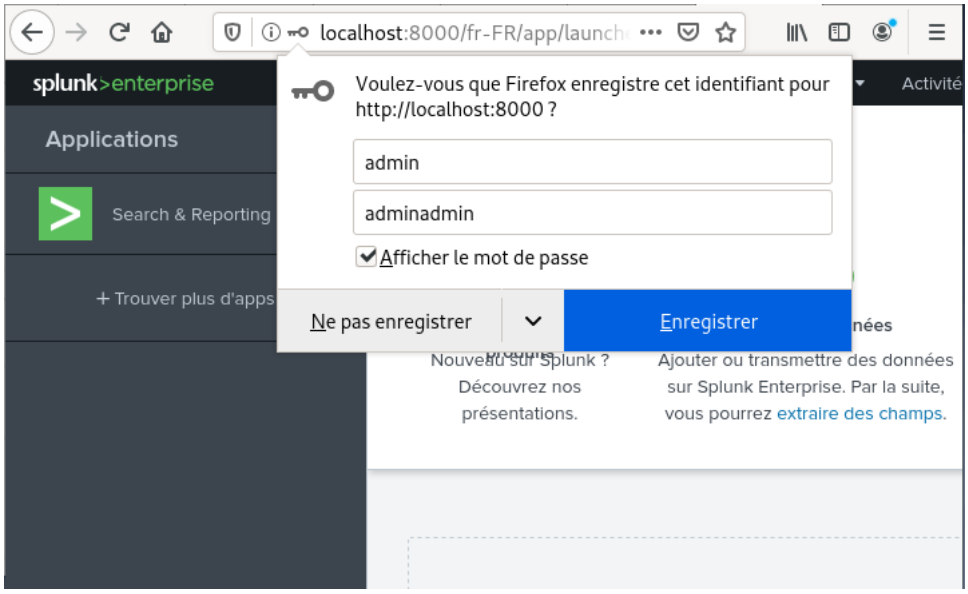
```
root@keycloak:/home/laurent# mkdir /opt/splunk
root@keycloak:/home/laurent# wget -O splunk-8.1.0.1-24fd52428b5a-Linux-x86_64.tg
z 'https://www.splunk.com/bin/splunk/DownloadActivityServlet?architecture=x86_64
&platform=linux&version=8.1.0.1&product=splunk&filename=splunk-8.1.0.1-24fd52428
b5a-Linux-x86_64.tgz&wget=true'
--2020-11-22 21:31:37--  https://www.splunk.com/bin/splunk/DownloadActivityServl
et?architecture=x86_64&platform=linux&version=8.1.0.1&product=splunk&filename=sp
lunk-8.1.0.1-24fd52428b5a-Linux-x86_64.tgz&wget=true
Résolution de www.splunk.com (www.splunk.com)… 95.100.252.147, 95.100.252.137
Connexion à www.splunk.com (www.splunk.com)|95.100.252.147|:443… connecté.
requête HTTP transmise, en attente de la réponse… 302 Moved Temporarily
Emplacement : https://download.splunk.com/products/splunk/releases/8.1.0.1/linux
/splunk-8.1.0.1-24fd52428b5a-Linux-x86_64.tgz [suivant]
--2020-11-22 21:31:38--  https://download.splunk.com/products/splunk/releases/8.
1.0.1/linux/splunk-8.1.0.1-24fd52428b5a-Linux-x86_64.tgz
Résolution de download.splunk.com (download.splunk.com)… 143.204.192.26, 143.204
.192.66, 143.204.192.77, ...
Connexion à download.splunk.com (download.splunk.com)|143.204.192.26|:443… conne
cté.
requête HTTP transmise, en attente de la réponse… 200 OK
Taille : 508221657 (485M) [application/x-gzip]
Sauvegarde en : « splunk-8.1.0.1-24fd52428b5a-Linux-x86_64.tgz »

        splunk-8.1.  39%[======>             ] 191,51M  5,48MB/s    tps 60s
```

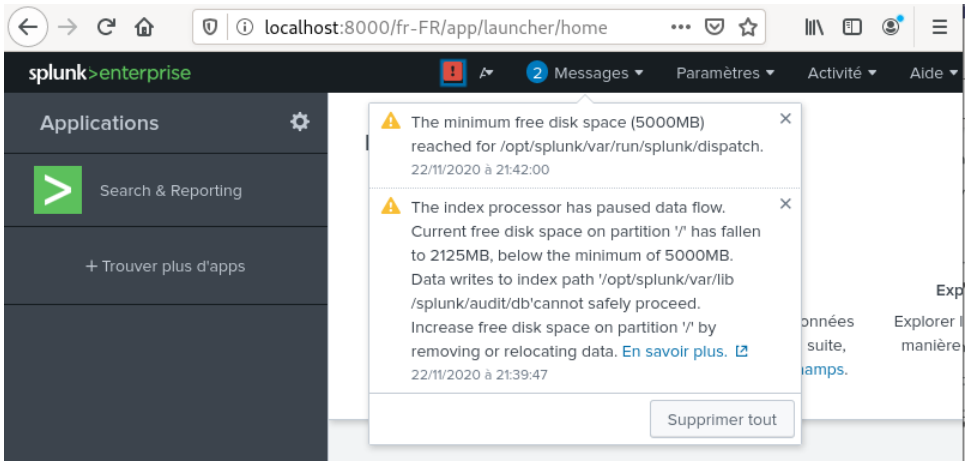
```
root@keycloak:/opt# tar -xvf splunk-8.1.0.1-24fd52428b5a-Linux-x86_64.tgz
```

```
root@keycloak:/opt/splunk/bin# ./splunk start

SPLUNK GENERAL TERMS

Last updated: February 13, 2020

These Splunk General Terms ("General Terms") between
Splunk Inc., a Delaware corporation, with its principal place
of business at 270 Brannan Street, San Francisco,
California 94107, U.S.A ("Splunk" or "we" or "us" or "our")
and you ("Customer" or "you" or "your") apply to the
purchase of licenses and subscriptions for Splunk's
Offerings. By clicking on the appropriate button, or by
downloading, installing, accessing or using the Offerings,
you agree to these General Terms. If you are entering into
these General Terms on behalf of Customer, you represent
that you have the authority to bind Customer. If you do not
agree to these General Terms, or if you are not authorized
to accept the General Terms on behalf of the Customer, do
not download, install, access, or use any of the Offerings.
```

```
Do you agree with this license? [y/n]: y

This appears to be your first time running this version of Splunk.

Splunk software must create an administrator account during startup. Otherwise,
you cannot log in.
Create credentials for the administrator account.
Characters do not appear on the screen when you type in credentials.

Please enter an administrator username: admin
Password must contain at least:
   * 8 total printable ASCII character(s).
Please enter a new password:
Please confirm new password:
Copying '/opt/splunk/etc/openldap/ldap.conf.default' to '/opt/splunk/etc/openlda
p/ldap.conf'.
Generating RSA private key, 2048 bit long modulus
.....................................................................................
.........+++++
.........+++++
e is 65537 (0x10001)
writing RSA key
```

```
If you get stuck, we're here to help.
Look for answers here: http://docs.splunk.com

The Splunk web interface is at http://keycloak:8000

root@keycloak:/opt/splunk/bin#
```

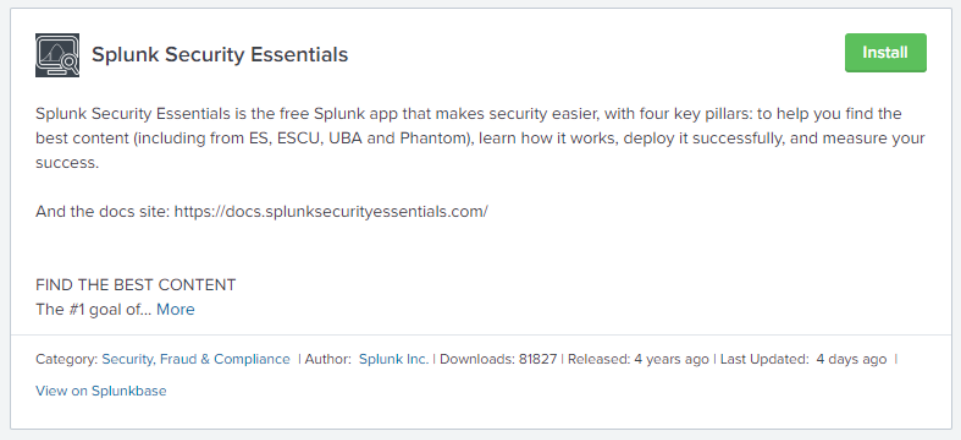




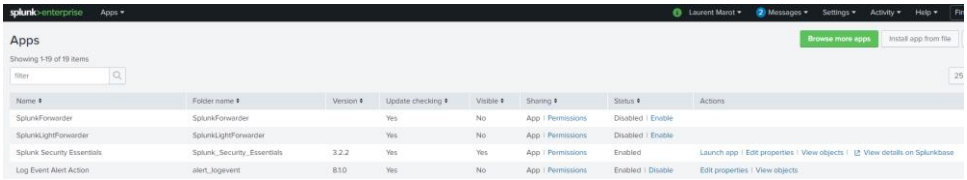
II] Installation d'une APP orientée "sécurité"

De base SPLUNK Enterprise n'est pas une solution de SIEM à proprement parler. La solution SIEM s'appelle "Enterprise sécurité" et est distribuée de manière payante. Et il faut justifier de la certification "Architecte SPLUNK" pour l'utiliser.

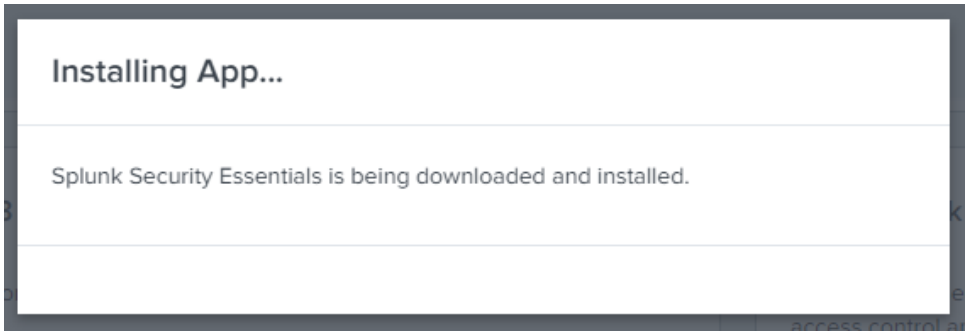
Malgré tout si on veut une petite appli didactique sympa orientée sécurité, nous avons "Splunk Security Essential "

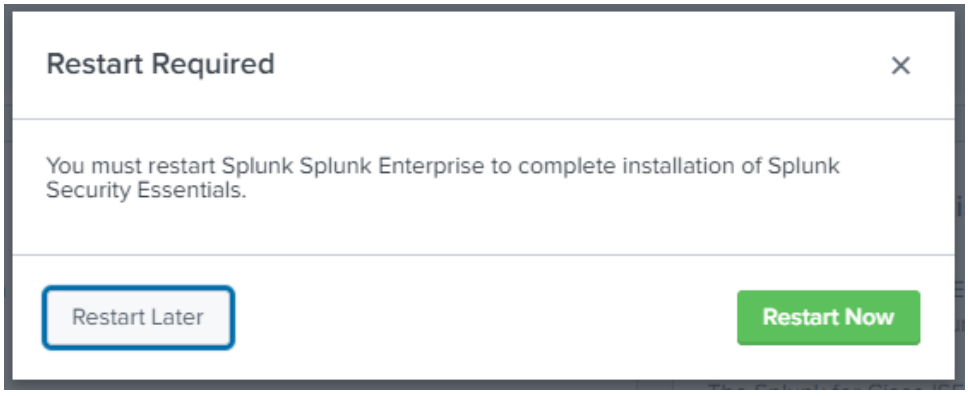


Découvrons au travers de cet exemple ce qu'est une APP Splunk.

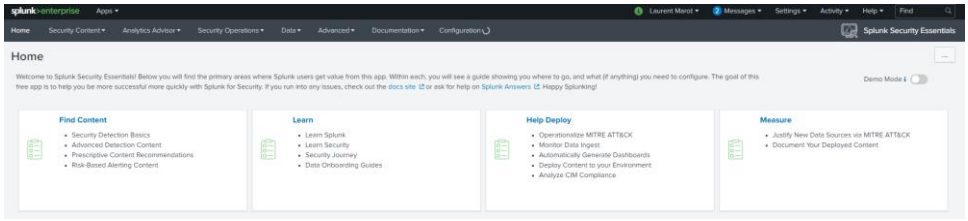


A la base, une APP c'est une petite tgz de fichier de scripting/coding/configuration qui vous permet de construire automatiquement tout un tas de recherches, alertes, rapports sur une source de logs particulière. Le code embarqué dans votre APP s'exécute avec les droits de votre utilisateurs SPLUNK alors méfiance.





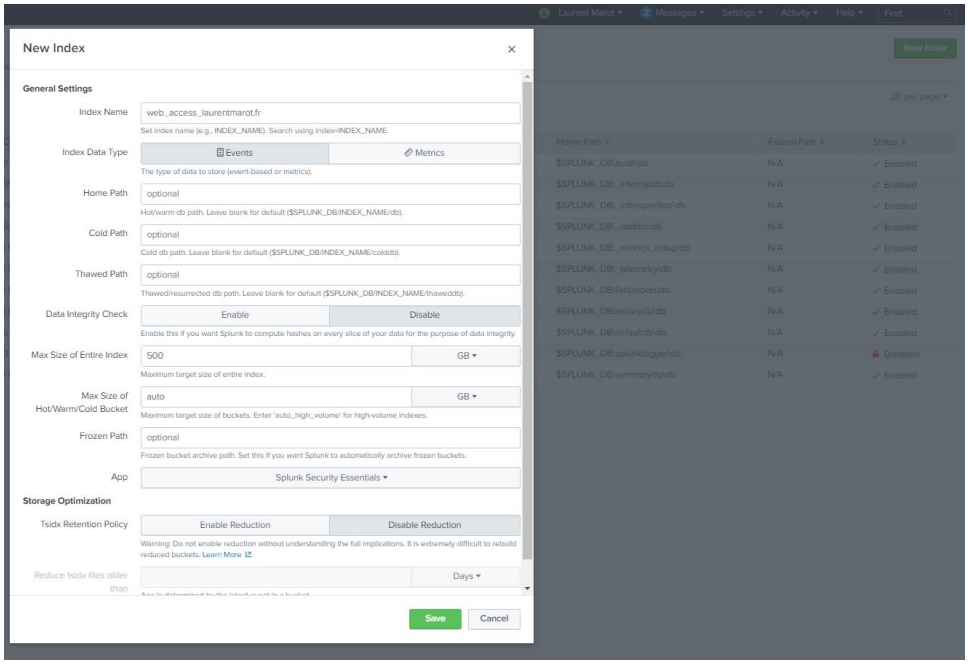
Larguez vos COD et vos World of Tanks, avec ça vous avez de quoi jouer pendant toutes vos vacances de Noël.



III] Playing with data

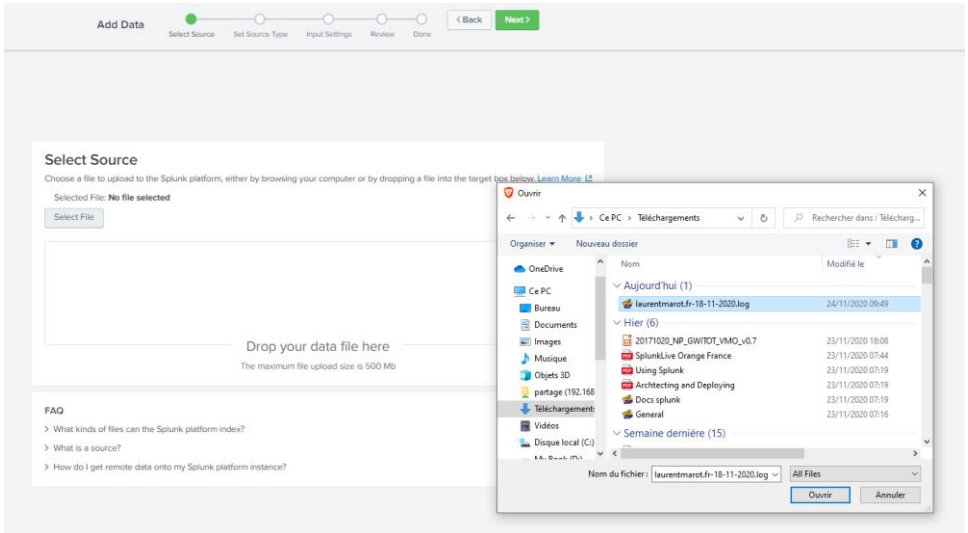
1 – acces web

Créons un index dédié :



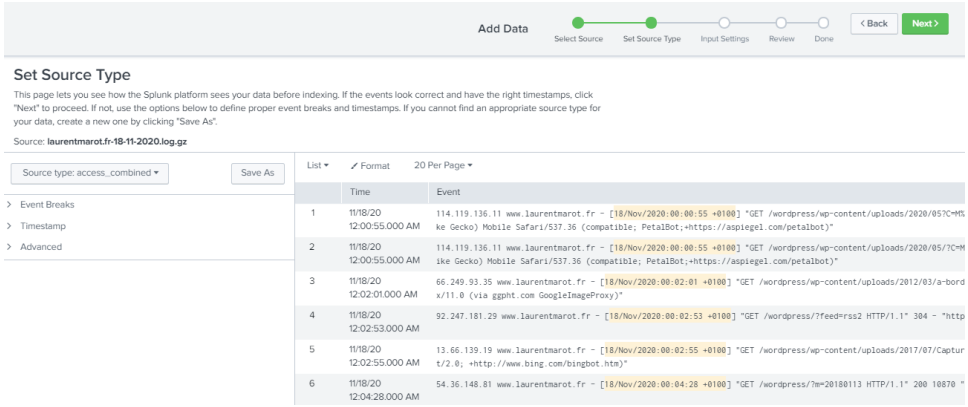
Attention d'avoir bien à l'esprit une convention de nommage explicite.

Et alimentons manuellement (dans un premier temps … ou simplement parce que les contraintes d'hébergement ne permettent pas de faire autrement).

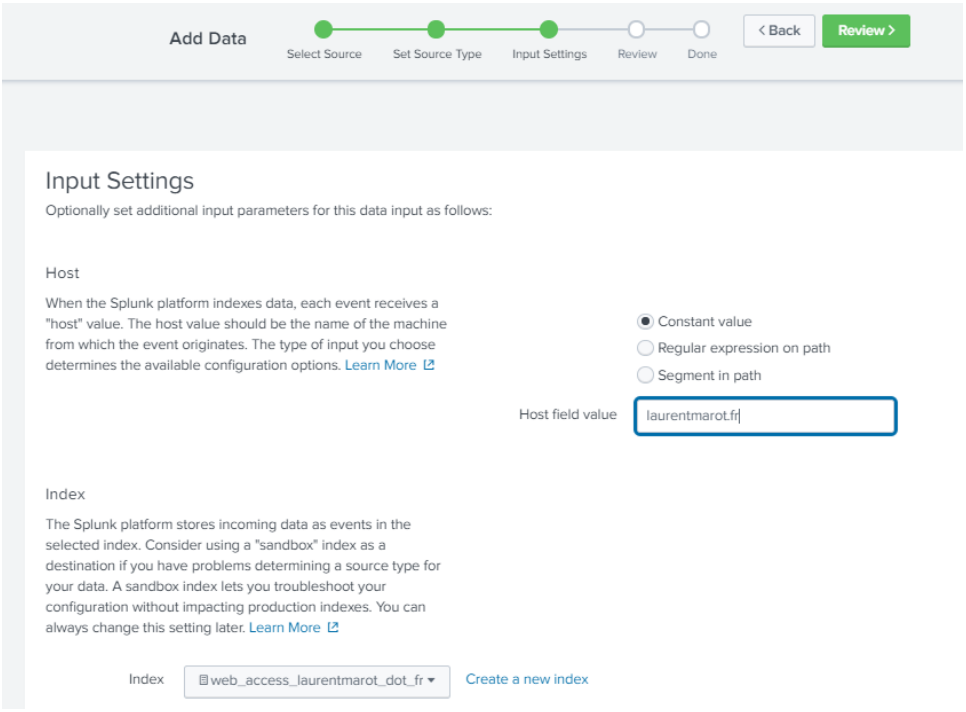


Procédure habituelle d'upload de fichiers (pour le coup logs Apache)

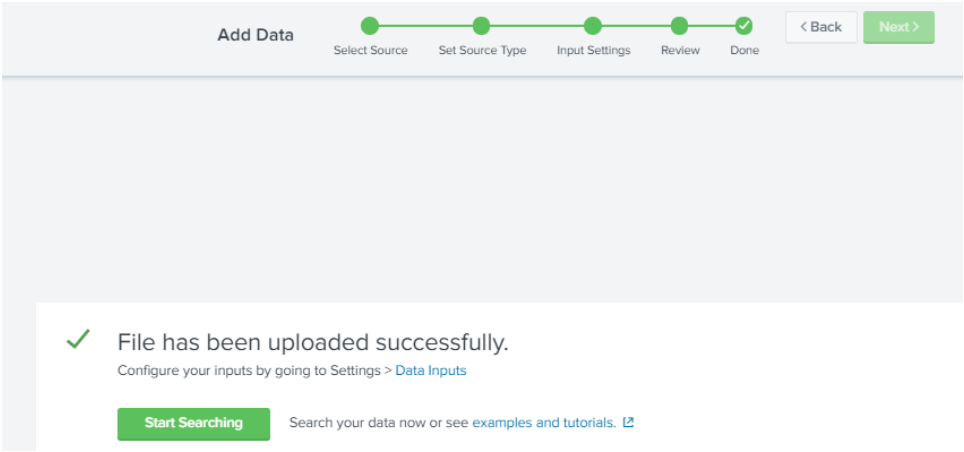
Comme le fichier est une archive gz, vous n'obtiendrez pas la visualisation des lignes de logs dans la première fenêtre du wizzard.



Le moteur Splunk détecte automatiquement que c'est bien un source-type "access_combined".

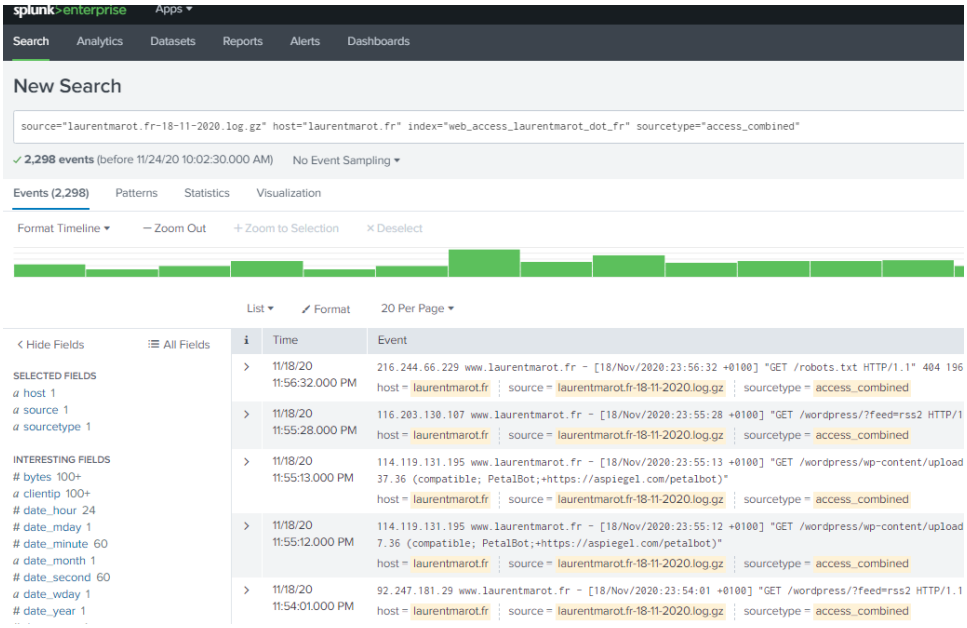


N'oubliez pas de sélectionner le bon index (et éventuellement changer le "host field value".



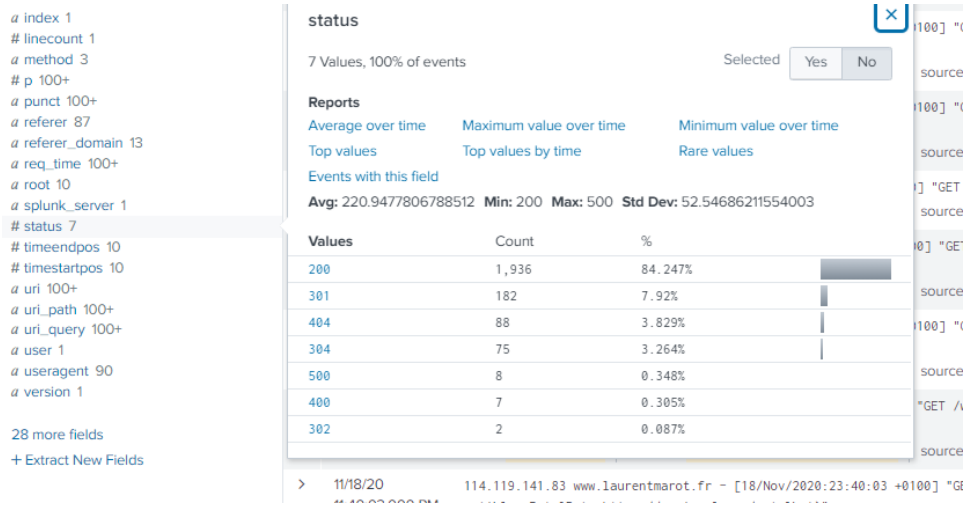
Tutto va bene !

Nous pouvons commencer nos opérations de recherche



Les champs significatifs sont reconnus(clientip, method, referer, …). Nous voilà parés pour faire un peu d'analytique et un peu de sécu.

Intéressons-nous au champs "status"



De deux choses l'une concernant les erreurs 404 : soit elles traduisent une mauvaise configuration de notre site (ressources absentes) soit ce sont des tentatives de découverte.

source="laurentmarot.fr-18-11-2020.log.gz"

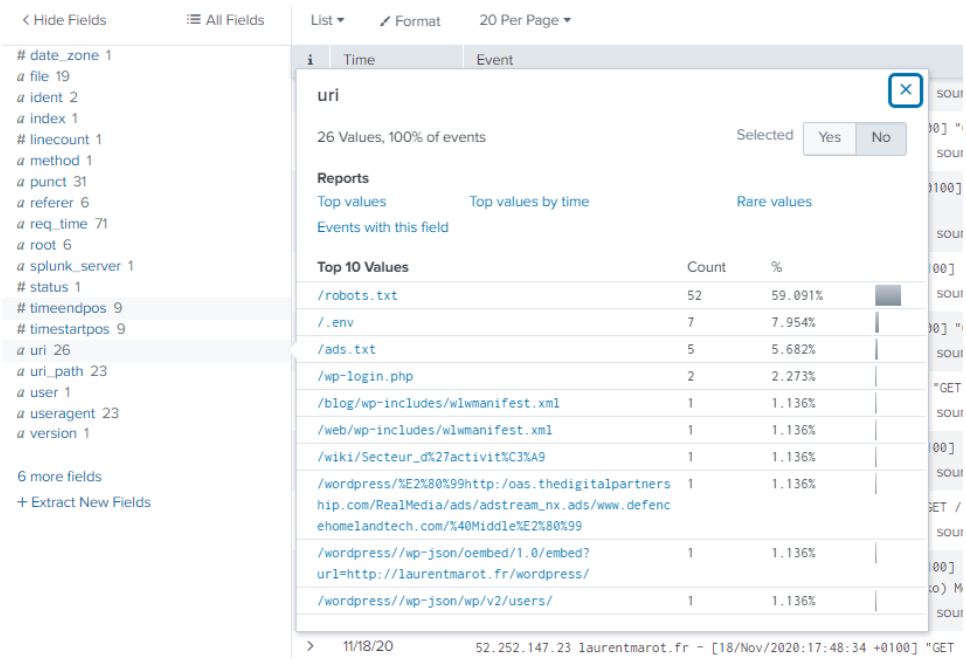
host="laurentmarot.fr"

index="web_access_laurentmarot_dot_fr"

sourcetype="access_combined"

status=404

Pour notre cas, 50 % des erreurs sont liées à l'absence du fichier robot.txt (consignes d'indexation aux moteurs de recherche). Mais qui gère ce site de daube ?

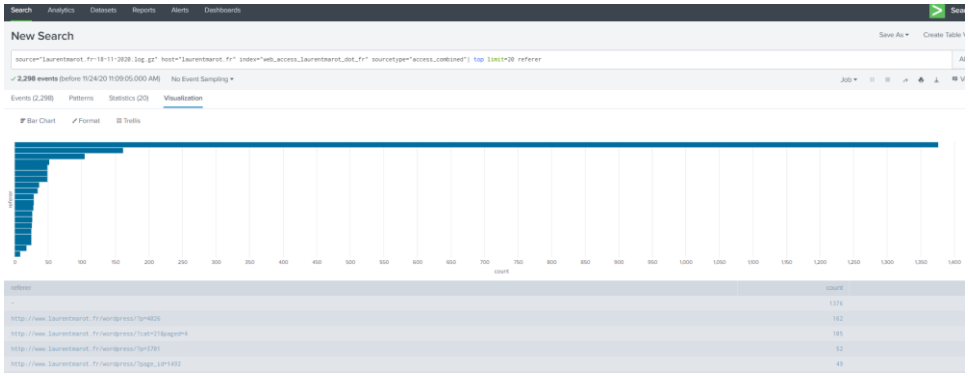


Je m'empresse donc d'ajouter un "robot.txt"

Quand on fait de l'analytique la question primordiale est : "qui vient sur mon site ?"

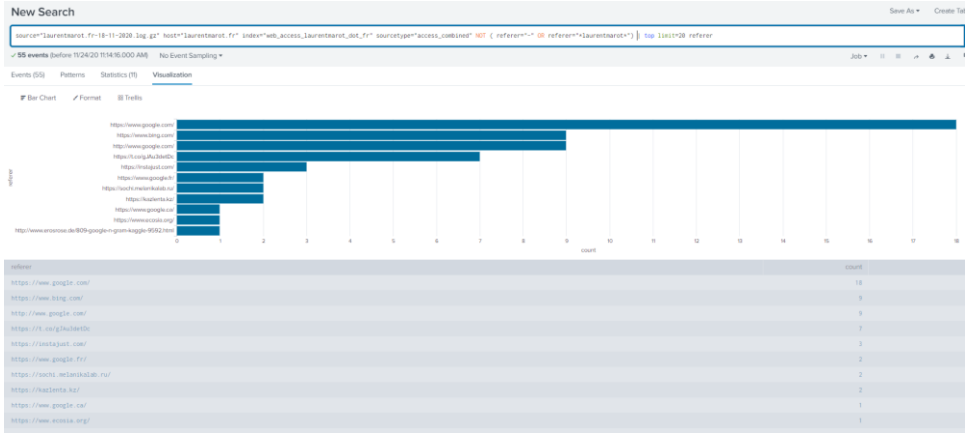
Et en particulier d'où viennent ces visites. Le champ "referer" est très important :

source="laurentmarot.fr-18-11-2020.log.gz" host="laurentmarot.fr" index="web_access_laurentmarot_dot_fr" sourcetype="access_combined"| top limit=20 referer



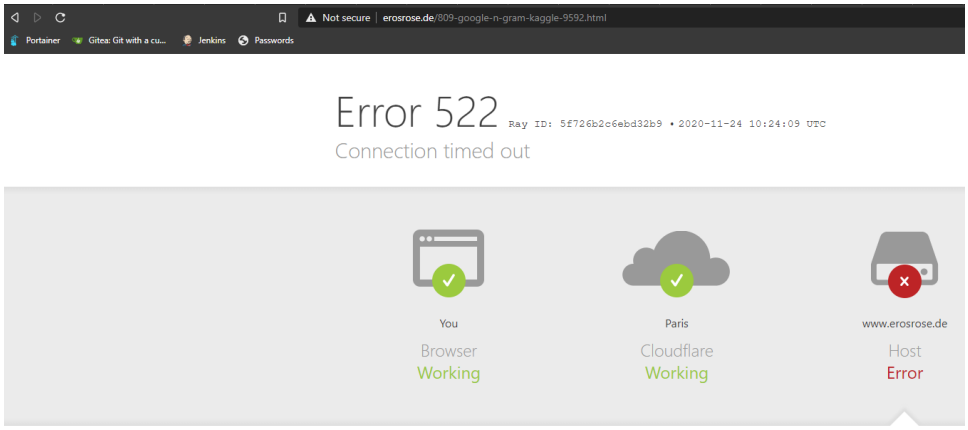
Naturellement, la liste est polluée par tous les enregistrements qui n'ont pas de referer (referer = -) ou dont la source est le site, lui-même.

Il est donc important de faire un peu de ménage...



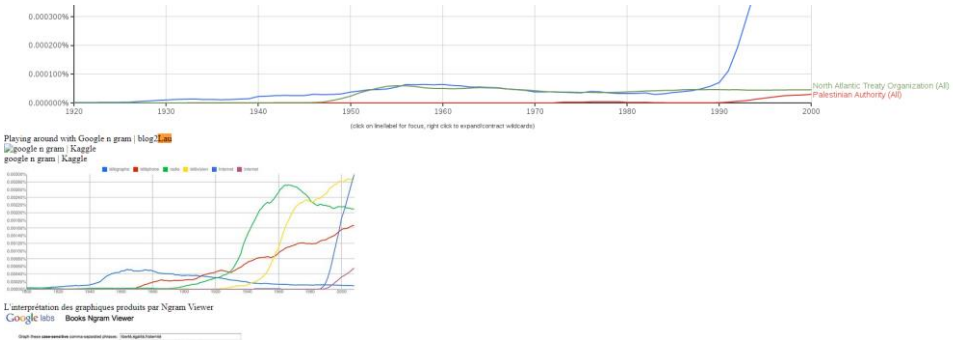
Et de se rendre compte qu'il va falloir vendre du lien "entrant" et payer une fortune à des experts en carton pour faire du SEO.

Un truc m'inquiète :



Le site semble down, je vois donc voir la version en cache.

https://webcache.googleusercontent.com/search?q=cache:aYDa98shNkwJ:www.erosrose.de/809-google-n-gram-kaggle-9592.html+&cd=1&hl=fr&ct=clnk&gl=fr

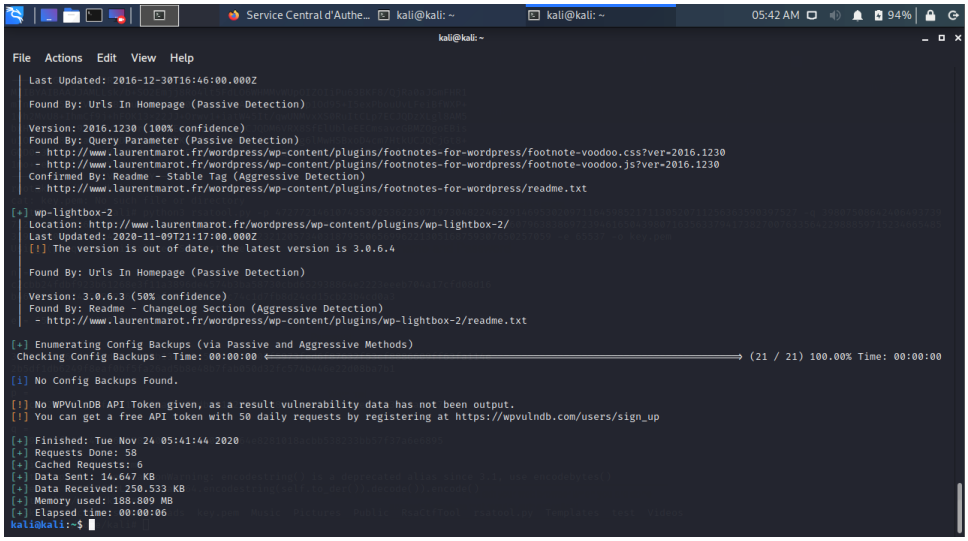


Pas grand-chose non plus sur webarchive, j'abandonne

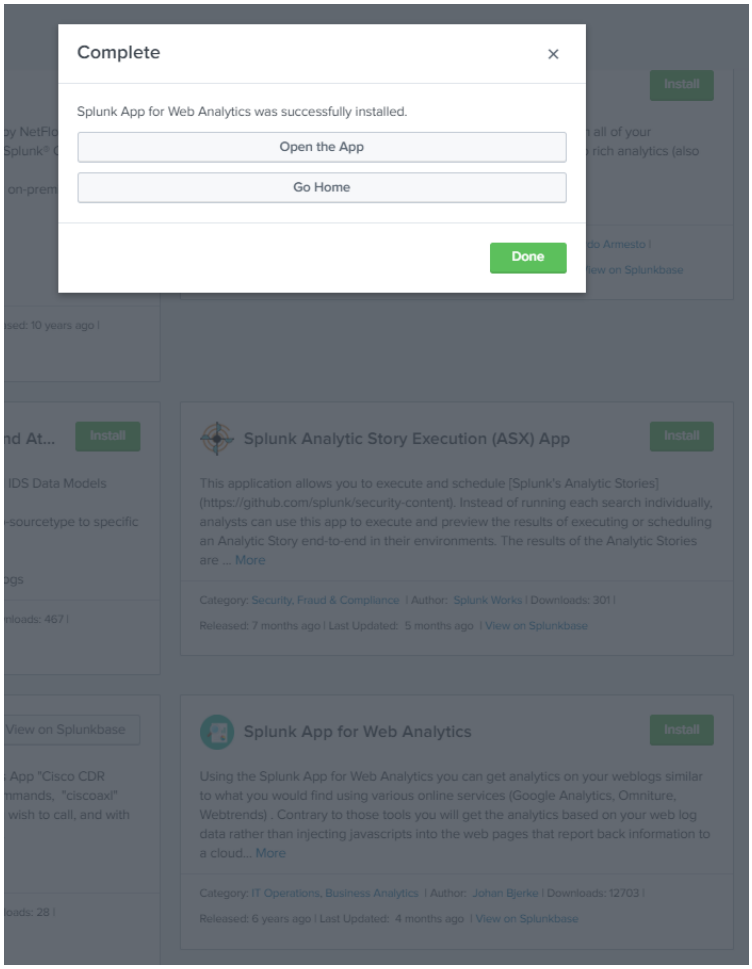
Je ne me souvenais pas davoir fait un post sur Google n gramm

http://www.laurentmarot.fr/wordpress/?s=google+n+gram&x=23&y=14

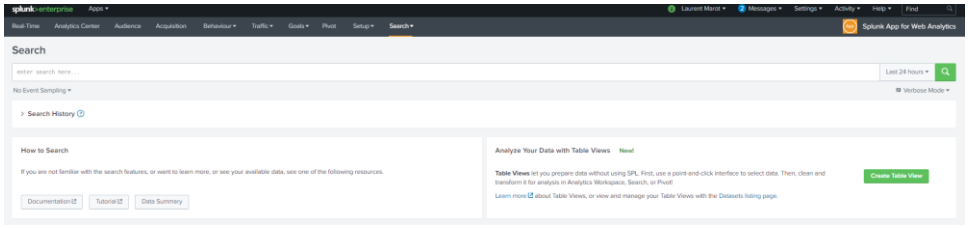
Pour faire un peu de bruit sur le site, nous allons jouer avec WPScan

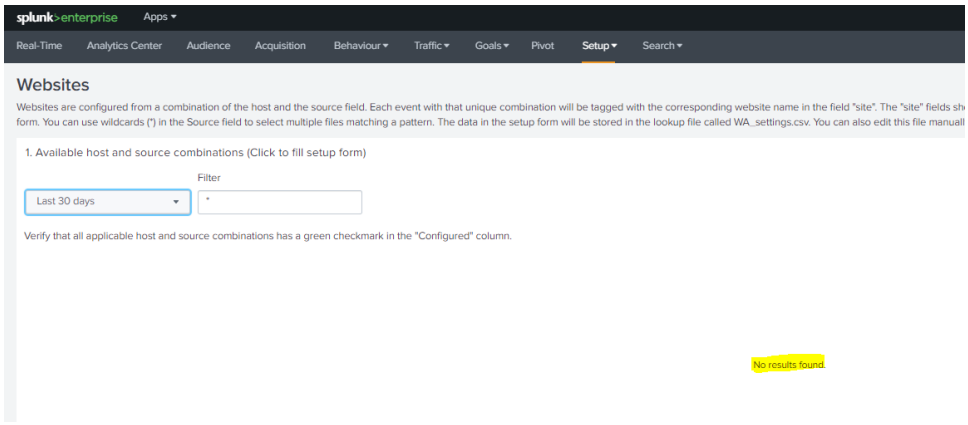


Les quelques requêtes faites précédemment montrent qu'avec un peu d'habitude et de patience on peut faire parler nos logs et savoir tout ce qui s'est passé sur notre serveur web. On peut ainsi créer manuellement toutes nos alertes et tableaux de bord. Pour procéder de manière plus performante, il existe un certain nombre d'applications qui vont faire le job à notre place.

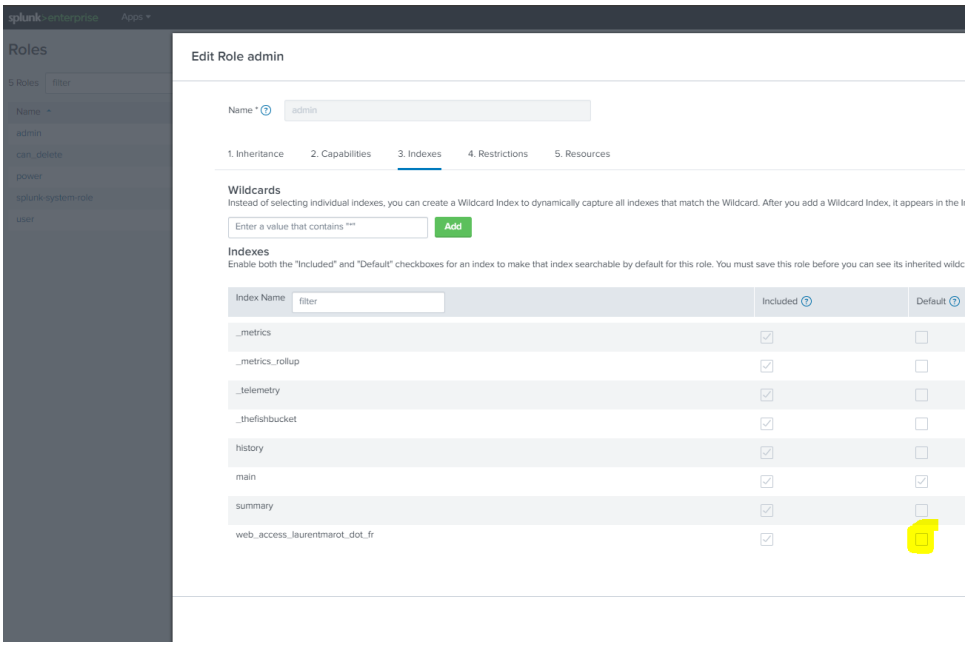


Un peu de configuration...





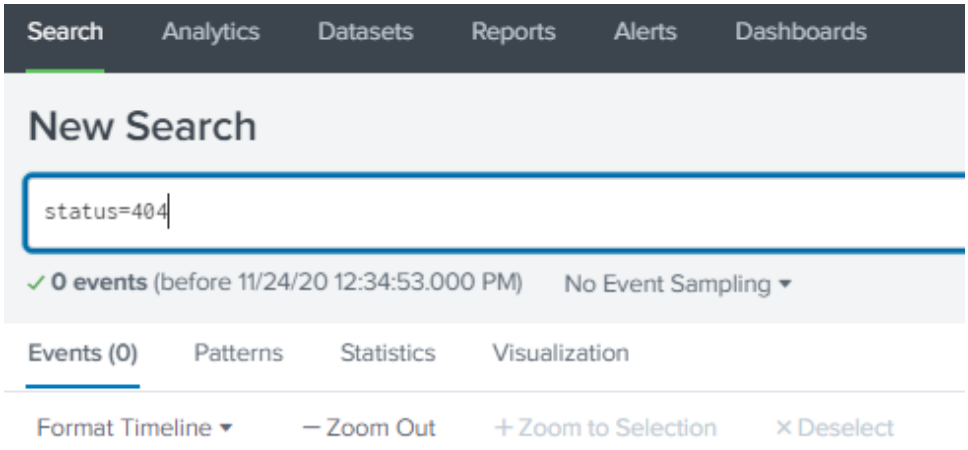
Très inquiétante information ? Non, juste un problème de droits pas défaut sur un index nouvellement créé.



Bongo, notre index créé pour stocker les logs d'accès de noter serveur web n'est pas interrogeable pas défaut, il faut donner explicitement son nom.

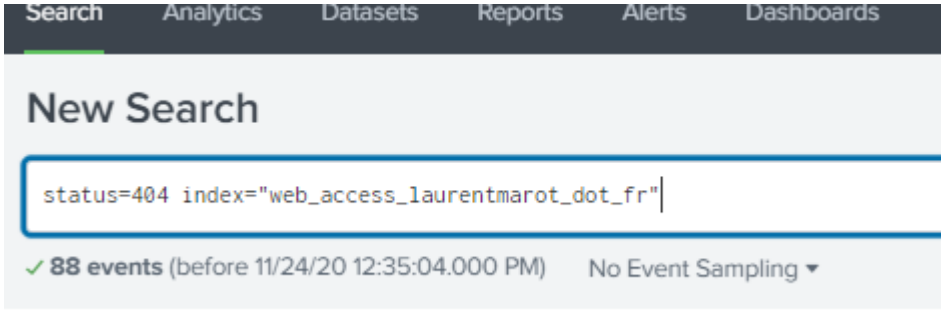
Pour vous en assurer, faites les requêtes suivantes :

status=404



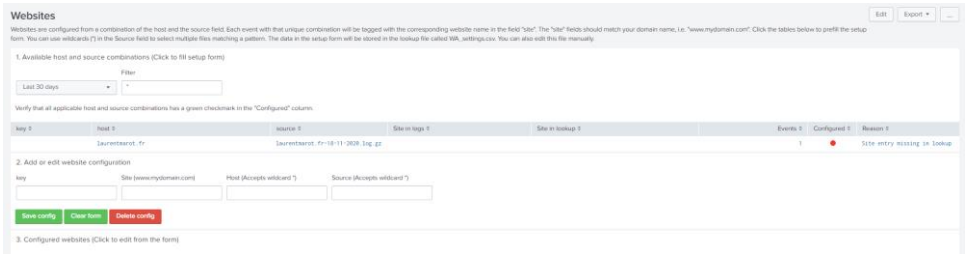
Vs

status=404 index="web_access_laurentmarot_dot_fr"



Pour que l'exemple soit probant, attention de placer le sélecteur de temps sur "All Times"

Et donc, après avoir donné les droits d'accès sur l'index, cela fonctionne mieux :



Sauvons la config

Websites

Websites are configured from a combination of the host and the source field. Each event with that unique combination will be tagged with the corresponding website name in the fiel
form. You can use wildcards (*) in the Source field to select multiple files matching a pattern. The data in the setup form will be stored in the lookup file called WA_settings.csv. You ca

1. Available host and source combinations (Click to fill setup form)

Filter

Last 30 days

*

Verify that all applicable host and source combinations has a green checkmark in the "Configured" column.

key	host	source	Site in logs
5fbcf0d30467000078003fe1	laurentmarot.fr	laurentmarot.fr-18-11-2020.log.gz	www.laurentmarot.fr

2. Add or edit website configuration

key	Site (www.mydomain.com)	Host (Accepts wildcard *)	Source (Accepts wildcard *)
	www.laurentmarot.fr	laurentmarot.fr	laurentmarot.fr-18-11-2020.log.gz

Save config | Clear form | Delete config

3. Configured websites (Click to edit from the form)

	Site	Host
1	www.laurentmarot.fr	laurentmarot.fr

4. Migrating config from v2.1 (csv) to >v2.1 (KV Store) - OPTIONAL

Below are the configs currently in the WA_settings.csv file. If you want to migrate this config to the KV store, click the button at the bottom. You only need to do this once.

Real-Time | Analytics Center | Audience | Acquisition | Behaviour | Traffic | Goals | Pivot | Setup | Search

Generate user sessions

Last 30 days

✓ **113 events** (10/25/20 12:00:00.000 AM to 11/24/20 12:40:06.000 PM)

113 results | 20 per page

acceleration	datamodel_update_time	count	_time	site	user	http_se
			2020-11-18 00:17:23	www.laurentmarot.fr	7653af1d34f8faad3a35c9bdd3664e25	0655c31
			2020-11-18 00:28:59	www.laurentmarot.fr	96d4d5cbfe66cda570441ca0c1fdb6bb	1ba94f7
			2020-11-18 00:36:32	www.laurentmarot.fr	898cd7a9344baf15d4d8940f90a69865	d9b700a
			2020-11-18 00:45:37	www.laurentmarot.fr	63a67369142c0a60d8e1e9b50170aa65	042dd87
			2020-11-18 00:57:33	www.laurentmarot.fr	5606f46ed123c302c64e62d276023a4a	b0accd4
			2020-11-18 01:05:42	www.laurentmarot.fr	829a2a82704072db7428ba6be053c945	a552972
			2020-11-18 01:28:41	www.laurentmarot.fr	c7860cc3fd06c7de7c6900edfbdb57aa	532b0de
			2020-11-18 01:29:01	www.laurentmarot.fr	24e954e4e9919a05f7b22b0225150589	565d0cc
			2020-11-18 02:01:59	www.laurentmarot.fr	15594fa6b238b6c46d23bad12e044965	96cff0b

Real-Time | Analytics Center | Audience | Acquisition | Behaviour | Traffic | Goals

Generate pages

Last 30 days

✓ **266 events** (10/25/20 12:00:00.000 AM to 11/24/20 12:40:35.000 PM)

100 results | 20 per page

site	http_request
www.laurentmarot.fr	/
www.laurentmarot.fr	/wordpress/?feed=rss2&cat=21
www.laurentmarot.fr	/wordpress/?p=4026
www.laurentmarot.fr	/wordpress/?feed=rss2
www.laurentmarot.fr	/wordpress/wp-content/themes/gazette/thumb.php?src=http://www.la
www.laurentmarot.fr	/wordpress/?page_id=2
www.laurentmarot.fr	/wordpress/?p=2364
www.laurentmarot.fr	/wordpress/?p=2263
www.laurentmarot.fr	/wordpress/?cat=21&paged=4
www.laurentmarot.fr	/wordpress/?p=3701
www.laurentmarot.fr	/wordpress/?cat=39
www.laurentmarot.fr	/index20200224.html
www.laurentmarot.fr	/wordpress/?page_id=1492

Et voilà de jolis tableaux de bord