

shinken-scheduler-export-data - export des données du Scheduler

Sommaire

- Intérêt global de la commande
- Fonctionnement général de la commande
 - Récupération des données des checks auprès des schedulers
 - Options de la commande
 - Options de la commande pour les extractions de types sizing-pollers et identify-most-consuming-checks
 - Résultat du lancement sur le terminal
 - Le retour du terminal
 - CSV des données
 - Contenu du fichier csv généré
 - Chargement et analyse des données
 - Chargement du fichier csv
 - Analyse des données
- Erreurs possibles
 - La commande est lancée depuis un serveur autre que l'Arbiter
 - Problème de paramétrage du mot de passe d'accès (`daemon__export_data__password` dans `/etc/shinken/shinken.cfg`)
 - Le mot de passe est vide ou non configuré dans le fichier `/etc/shinken/shinken.cfg`
 - Le mot de passe fourni par l'utilisateur est incorrect
 - Problème de paramétrage de la désactivation de l'export des données des démons (`scheduler__export_data__enabled` parameter dans le `.cfg` du Scheduler)
 - Problèmes avec le réseaux ou l'état du Scheduler
 - Le Scheduler n'est pas accessible (éteint ou problème réseau)
 - Le Scheduler n'a pas encore reçu de configuration de l'Arbiter
 - Passer les erreurs et travailler avec des données partielles

Intérêt global de la commande

Sur les grands environnements, la majorité de la charge CPU va être consommée par les sondes de supervision. Quand elles sont nombreuses et qu'elles sont ordonnancées avec des intervalles de temps différents, il peut être difficile :

- d'extrapoler, par rapport au nombre d'hôtes déjà présents, de combien de CPUs sur ses Pollers une infrastructure à besoin.
- de savoir, sur une infrastructure, qu'elles sont les sondes les plus consommatrices en termes de CPU :
 - une fois identifiées, il est intéressant de passer du temps pour optimiser ces sondes afin de diminuer le nombre de CPU nécessaires.

C'est pour répondre à ces problématiques que nous avons créé la commande **shinken-scheduler-export-data**.

Vous trouverez des exemples d'analyses de résultat de la commande dans les pages suivantes :

- Vous avez un environnement avec un certain nombre d'hôtes, vous souhaitez dimensionner vos Pollers si vous rajoutez 10, 20 ou 50% d'hôtes en plus :
 - dans ce cas, il faut se référer à la page [Dimensionnement des Pollers à l'aide de la commande shinken-scheduler-export-data](#).
- Si vous souhaitez diminuer le nombre de CPU de vos Pollers en identifiant, et optimisant, les sondes les plus consommatrices :
 - dans ce cas, il faut se référer à la page [Identification des checks les plus consommateurs à l'aide de la commande shinken-scheduler-export-data](#).

Cette commande interroge les Schedulers, car ce sont eux qui centralisent les informations de temps d'exécution des sondes de supervision exécutées par les Pollers.

- Le fichier de résultat est un fichier `.csv` qui contient une ligne par check exécuté par les Pollers, avec suivant les options des données différentes.
 - Ces données seront les données constatées par les Pollers.
- Vous pouvez aussi simuler la génération des lignes de check sur une période.
 - L'intérêt est que si certains checks sont exécutés toutes les heures, ils n'auront pas le même poids dans vos besoins en ressource qu'un check exécuté toutes les minutes.

Fonctionnement général de la commande

Récupération des données des checks auprès des schedulers

La commande "shinken-scheduler-export-data" est seulement utilisable depuis le serveur de l'**Arbiter**.

- En effet, elle se connecte sur tous les Schedulers, et c'est le seul démon habilité à cela.

La commande va lire la configuration depuis /etc/shinken et va pouvoir se connecter sur tous les Schedulers qui sont définis dans l'architecture. Le fichier de résultat peut être :

- soit ANONYME (*aucun nom d'hôte, royaume, etc, ne sera présent dans le fichier généré, voir l'option **--anonymous***),
 - L'extraction sera anonymisée (*seule un hash des noms sera lisible*)
- soit NORMAL, donc avec les noms des éléments (*par défaut*)
 - Un mot de passe vous sera systématiquement demandé pour générer ce fichier
 - Le mot de passe est celui défini dans le fichier de configuration de Shinken (/etc/shinken/shinken.cfg) dans la clé : **daemon__export_data__password** (voir la page [Fichier de configuration \(shinken.cfg \)](#)).

Il est possible de désactiver l'extraction pour un Scheduler via l'option définie dans son fichier de configuration avec le paramètre : **scheduler__export_data__enabled** (voir la page [Le Scheduler](#)).

Options de la commande

La commande accepte les options suivantes :

Nom	Type	Unité	Défaut	Commentaire
<code>--export-type</code>	Texte	--	--	Cette option permet de choisir quel type de donnée sera exportées. Les valeurs possibles de cette option sont : • sizing-pollers: Export les données pour le dimensionnement des Poller. • identify-most-consuming-checks: Export les données pour identifier les sondes les plus consommatrices. <code>--export-type=sizing-pollers</code>
<code>--anonymous</code>	--	--	--	Cette option permet d'avoir un export anonyme, c'est-à-dire que les noms des éléments ne seront pas présents. • Cela vous permettra d'exporter les données dans les situations où la confidentialité est nécessaire. • Par exemple, mettre à disposition ces données a votre support Shinken sans que les noms de vos machines ne soient diffusés <code>--anonymous</code>
<code>--scheduler-timeout</code>	Entier	sec	10	Permet de configurer le temps d'attente acceptée sur les appels des Schedulers avant de retourner une erreur. <code>--scheduler--timeout=30</code>
<code>--realm-filter</code>	Texte	--	--	Prends le nom du royaume que l'on souhaite utiliser comme filtre afin de n'avoir que des Schedulers de ce royaume dans notre extraction de données. (<i>on ne peut mettre qu'un seul royaume dans le filtre</i>) <code>--realm-filter=France</code>
<code>--dont-stop-on-scheduler-error</code>	Booléen	--	--	Si activé, la commande ne s'arrêtera plus si un ou plusieurs Scheduler est éteint/mal configuré, ou s'il ne répond pas (<i>timeout</i>). A noter que dans ce cas les données extraites risquent d'être partielles avec des éléments manquants. De plus, les Schedulers spare inactifs sont automatiquement passés, il n'est pas nécessaire d'utiliser cette option pour ces derniers. <code>--dont-stop-on-scheduler-error</code>

--debug	--	--	--	Option permettant d'avoir plus d'affichage sur le fonctionnement interne de la commande. Pour le support uniquement. <pre>--debug</pre>
---------	----	----	----	--

Options de la commande pour les extractions de types sizing-pollers et identify-most-consuming-checks

Nom	Type	Unité	Défaut	Commentaire
--simulate-scheduling-for-X-seconds	Entier	sec	0	Cette option permet, de simuler une activité normale des checks, en rajoutant leurs données après chaque intervalle de vérification spécifique à chaque check. - Une ligne de donnée sera donc ajoutée pour chaque exécution simulée. Par exemple: avec <code>simulate-scheduling-for-X-seconds=3600</code>, un check avec un intervalle de vérification à une minute générera 60 lignes.  Cette simulation ne prend pas en compte : - les périodes de vérification (<i>prévu pour une futur version</i>), - les intervalles de nouvelles tentatives de confirmations d'état (<i>retry_interval</i>) => Difficile de définir de manière pertinente combien d'incidents futurs pourraient apparaître et engendrer une activité supplémentaire sur les Pollers à cause de ces sur-vérifications de confirmation. Donc les valeurs données par cette simulation ne seront pas forcément parfaitement alignées avec la réalité du futur. <pre>--simulate-scheduling-for-X-seconds=3600</pre>

Résultat du lancement sur le terminal

Le lancement de la commande fera deux choses :

- un retour dans le terminal pour voir le retour global de la commande,
- la génération d'un fichier .csv avec l'ensemble des données.
 - Les lignes du fichier csv représente chaque occurrence d'exécution de chaque check.

Deux exemples seront fournis pour illustrer la documentation:

- Exemple 1 : Exécution sécurisée des données pour le dimensionnement des Pollers (`--export-type=sizing-pollers`) et sans simulation dans le futur.

```
shinken-scheduler-export-data --export-type=sizing-pollers
```

- Exemple 2 : Exécution des données pour identifier les sondes qui consomme le plus de CPU sur les Pollers (`--export-type=identify-most-consuming-checks`) avec une demande de simulation d'une heure supplémentaire (`--simulate-scheduling-for-X-seconds=3600`), de manière anonyme (`--anonymous`) pour ne pas avoir les noms des hôtes.

```
shinken-scheduler-export-data --export-type=identify-most-consuming-checks --simulate-scheduling-for-X-seconds=3600 --anonymous
```

Le retour du terminal

Le retour donne :

- la liste des Schedulers,
- l'adresse du fichier où le CSV a été généré,
- le nombre de lignes générées (*qui correspond aux nombres d'éléments et simulé si l'option `--simulate-scheduling-for-X-seconds` a été utilisé*)

Exemple 1

```
#shinken-scheduler-export-data --export-type=sizing-pollers
Enter export data password ("daemon_export_data_password" in file /etc/shinken/shinken.cfg):

Get check dumps from schedulers (export with names)

All >
- scheduler-spare           : SKIPPED (sleeping spare)
- scheduler-secondary       : OK
- scheduler-master          : OK
[ sizing-pollers ] Writing to CSV file: /tmp/dump-schedulers--sizing-pollers--no-simulate-scheduling--with-names--2023-04-24-10h04m05s.csv
Wrote 26 lines in 0.001s
```

Exemple 2

```
#shinken-scheduler-export-data --export-type=identify-most-consuming-checks --anonymous --simulate-scheduling-for-X-seconds=3600
Get check dumps from schedulers (anonymous dump)

All >
- scheduler-spare           : SKIPPED (sleeping spare)
- scheduler-secondary       : OK
- scheduler-master          : OK
[ identify-most-consuming-checks ] Writing to CSV file: /tmp/dump-schedulers--identify-most-consuming-checks--simulate-scheduling-for-3600-seconds--anonymous--2023-04-24-10h05m29s.csv
Wrote 530 (with 504 simulated scheduling) lines in 0.006s
```

À noter que dans l'exemple 2, le nombre de lignes est bien plus élevées à causes des nombreuses lignes de la simulation (530 au lieu de 26 dans notre exemple).

CSV des données

Elle va générer un fichier **csv** dans le répertoire **/tmp** du serveur dont le nom va dépendre des options utilisées :

- Exemple 1 :
/tmp/dump-schedulers--sizing-pollers--no-simulate-scheduling--with-names--2023-04-24-10h04m05s.csv
- Exemple 2 :
/tmp/dump-schedulers--identify-most-consuming-checks--simulate-scheduling-for-3600-seconds--anonymous--2023-04-24-10h05m29s.csv

Contenu du fichier csv généré

Dans le fichier CSV, on a les données suivantes :

host_uuid	UUID de l'hôte.
host_name	Nom de l'hôte (Vide si l'option --anonymous est défini dans la commande).
check_uuid	UUID du check.
check_name	Nom du check (Vide si l'option --anonymous est défini dans la commande); vide si la vérification porte sur un hôte
command_name	Nom de la commande (Vide si l'option --anonymous est défini dans la commande) .
command_name_anonymous_hash	Hash du nom pour l'identifier sur un export avec l'option --anonymous .
realm	Nom du royaume (Vide si l'option --anonymous est défini dans la commande) .
realm_anonymous_hash	Hash du nom du royaume pour l'identifier sur un export avec l'option --anonymous .
check_interval	Intervalle entre les vérifications.
retry_interval	Intervalle de nouvelles tentatives de confirmations d'état.

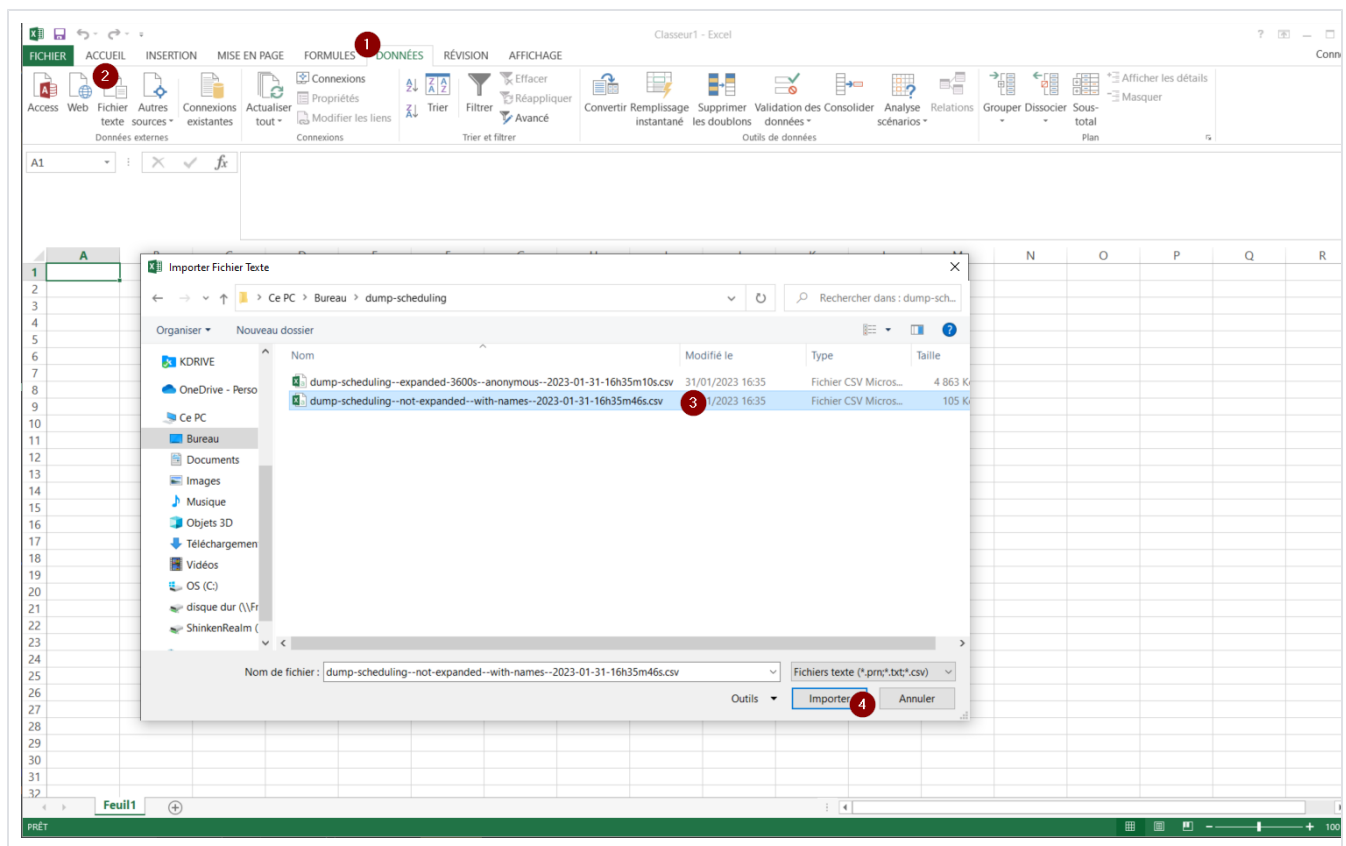
cpu_time	Temps CPU utilisé par le check : • Format chaîne de caractère représentant un nombre flottant (<i>float</i>) avec un "." • À bien faire attention quand on charge dans un Excel français, il faut changer le séparateur de "," à "."
scheduled_epoch	Seconde où le check est planifié (<i>en epoch, depuis le 1er janvier 1970</i>).
scheduled_date	Format YYYY-MM-DD hh:mm:ss
is_simulated	Booléen permettant de savoir si le check est issu d'une vraie exécution prévue dans le Scheduler, ou s'il est issu d'une simulation du check dans le futur (<i>via l'option --simulate-scheduling-for-X-seconds</i>)
last_reference_epoch	Seconde où le check s'est exécuté pour la dernière fois et que son état a été confirmé (" <i>HARD</i> ") (<i>en epoch, depuis le 1er janvier 1970</i>).

Chargement et analyse des données

Chargement du fichier csv

Les fichiers CSV générés peuvent être importés dans un tableur comme Excel.

- L'import dans ce dernier se fait simplement, dans les autres tableurs l'import suivra un mécanisme similaire.
- Depuis un tableur vide, il faut passer par **Données**Fichier texte.



Vu que le fichier a une première ligne avec la description des champs, il faut bien activer "Mes données ont des en-têtes".

Assistant Importation de texte - Étape 1 sur 3

L'Assistant Texte a déterminé que vos données sont de type Délimité.

Si ce choix vous convient, choisissez Suivant, sinon choisissez le type de données qui décrit le mieux vos données.

Type de données d'origine

Choisissez le type de fichier qui décrit le mieux vos données :

☒ Délimité - Des caractères tels que des virgules ou des tabulations séparent chaque champ.

☐ Largeur fixe - Les champs sont alignés en colonnes et séparés par des espaces.

Commencer l'importation à la ligne : Origine du fichier : MS-DOS (PC-8)

☒ Mes données ont des en-têtes.

Aperçu du fichier C:\Users\j.gabes\Desktop\dump-scheduling\dump-scheduling--not-expanded--with-names--2023-02-01-10h22m06s.csv.

1	"#host_uuid","host_name","host_name_anonymous_hash","check_uuid","check_name","check_name_anonymous_hash","c
2	"dfe65e40072b41b8950ed69d1329cc48","srv-15","anonymous-hash-0f628613ec","0f31a3424df44ae41053d5c72dbc7eda","
3	"1b7ce868df5547038603eec61616cae2","srv-7","anonymous-hash-eb645d6b77","24e86f31b7f133bda96ac03a67ff0f14","d
4	"8bc339daa0a31leda85d080027940ca8","srv-2","anonymous-hash-6fb60b8226","bb64373c94c329a13609798292c7db56","d
5	"8bcd91c8a0a31ledb99d080027940ca8","srv-8","anonymous-hash-35a25c2308","525e131883b1d7c7c30901fef78e992b","d
6	"8bd19548a0a31led8e0f080027940ca8","srv-10","anonymous-hash-a89f7681d1","bb64373c94c329a13609798292c7db56","

Annuler < Précédent Suivant > Terminer

Sur la phase suivante, il faut bien placer le séparateur sur "Virgule".

- L'Aperçu de données permet de s'apercevoir que la séparation des champs est bien prise en compte.

Assistant Importation de texte - Étape 2 sur 3

Cette étape vous permet de choisir les séparateurs contenus dans vos données. Vous pouvez voir les changements sur votre texte dans l'aperçu ci-dessous.

Séparateurs

☐ Tabulation

☐ Point-virgule

☒ Virgule

☐ Espace

☐ Autre :

☐ Interpréter des séparateurs identiques consécutifs comme uniques

Identificateur de texte :

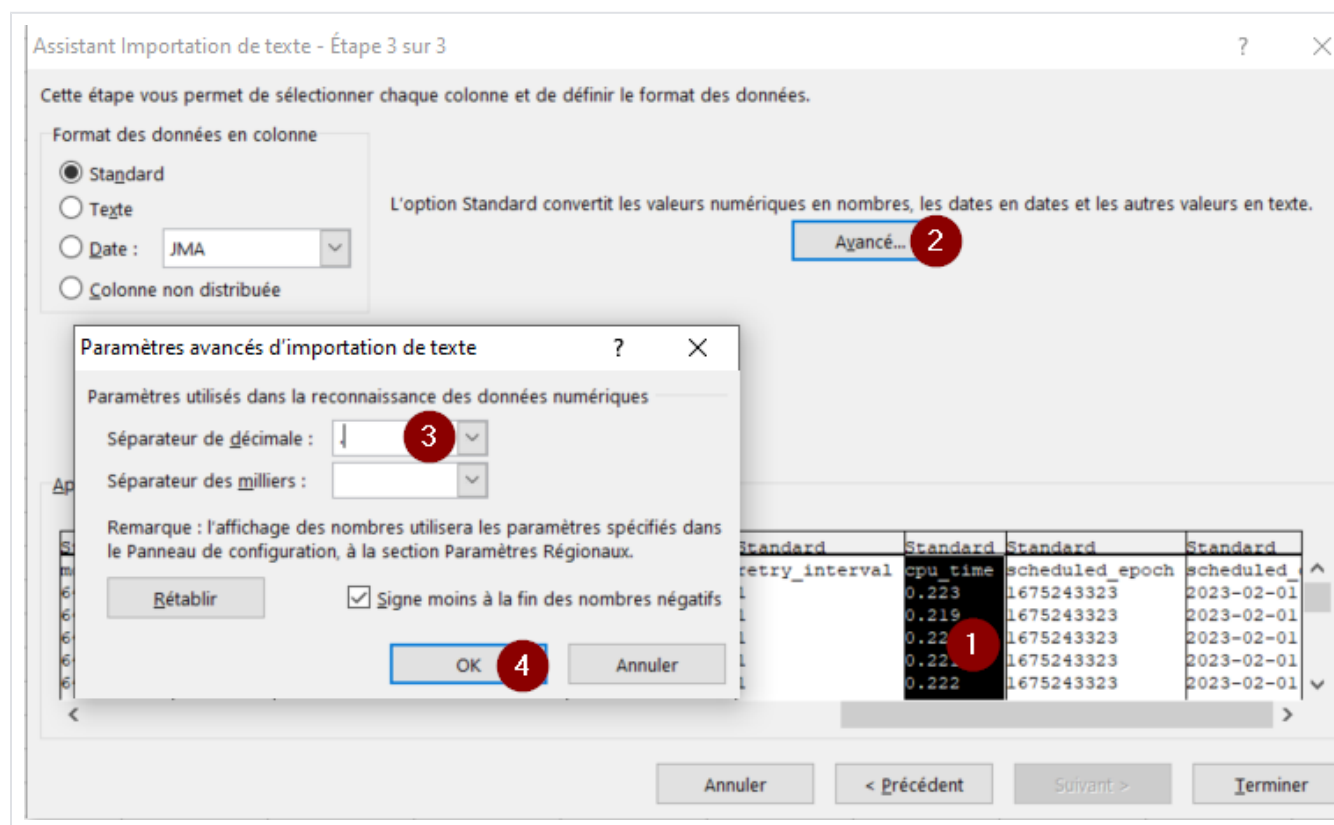
Aperçu de données

#host_uuid	host_name	host_name_anonymous_hash	check_uuid	check_name
dfe65e40072b41b8950ed69d1329cc48	srv-15	anonymous-hash-0f628613ec	0f31a3424df44ae41053d5c72dbc7eda	dump-che
1b7ce868df5547038603eec61616cae2	srv-7	anonymous-hash-eb645d6b77	24e86f31b7f133bda96ac03a67ff0f14	dump-che
8bc339daa0a31leda85d080027940ca8	srv-2	anonymous-hash-6fb60b8226	bb64373c94c329a13609798292c7db56	dump-che
8bcd91c8a0a31ledb99d080027940ca8	srv-8	anonymous-hash-35a25c2308	525e131883b1d7c7c30901fef78e992b	dump-che
8bd19548a0a31led8e0f080027940ca8	srv-10	anonymous-hash-a89f7681d1	bb64373c94c329a13609798292c7db56	dump-che

Annuler < Précédent Suivant > Terminer

Si la majorité des champs sont des formats simples (*chaines de caractères ou entier*), deux colonnes sont particulières et on doit indiquer à Excel comment les interpréter :

- **cpu_time** : Il faut lui indiquer via "Avancé" que le séparateur de décimale n'est pas "," mais "." (à faire que dans un Excel en Français, car dans le Excel en Anglais le séparateur par défaut est déjà le point).



- **scheduled_date** : Il faut lui indiquer que le champ est un format Date.

Assistant Importation de texte - Étape 3 sur 3

Cette étape vous permet de sélectionner chaque colonne et de définir le format des données.

Format des données en colonne

☐ Standard
☐ Texte
☒ Date : JMA 2
☐ Colonne non distribuée

L'option Standard convertit les valeurs numériques en nombres, les dates en dates et les autres valeurs en texte.

Avancé...

Aperçu de données

Standard	Standard	Standard	Standard	Standard	Standard	Standard	JMA
mous_hash	realm	realm_anonymous_hash	check_interval	retry_interval	cpu_time	scheduled_epoch	scheduled
6432587	All	anonymous-hash-6a72085653	1	1	0.223	1675243323	2023-02-01
6432587	All	anonymous-hash-6a72085653	1	1	0.219	1675243323	2023-02-01
6432587	customer	anonymous-hash-b39f008e31	1	1	0.221	1675243323	2023-02-01
6432587	customer	anonymous-hash-b39f008e31	1	1	0.221	1675243323	2023-02-01
6432587	customer	anonymous-hash-b39f008e31	1	1	0.222	1675243323	2023-02-01

3

Excel nous demande alors où charger ces données. Il est plus simple et lisible de le charger dans une **"Nouvelle feuille de calcul"** que l'on pourra renommer facilement.

Importer des données

Sélectionnez la méthode d'affichage de ces données dans votre classeur.

☒ Tableau
☐ Rapport de tableau croisé dynamique
☐ Graphique croisé dynamique
☐ Ne créer que la connexion

Insérer les données dans

☐ Feuille de calcul existante :
 =SAS1
☒ Nouvelle feuille de calcul 1

☐ Ajouter ces données au modèle de données

2

Voici un exemple de ce qu'on obtient finalement :

Analyse des données

L'analyse des données issues du script va être répartie en deux grandes catégories :

- Vous avez un environnement avec un certain nombre d'hôtes, vous souhaitez dimensionner vos Pollers si vous rajoutez 10, 20 ou 50% d'hôtes en plus :
 - Dans ce cas, il faut se référer à la page [Dimensionnement des Pollers à l'aide de la commande shinken-scheduler-export-data](#)
- Si vous souhaitez diminuer le nombre de CPU de vos Pollers en identifiant, et optimisant, les sondes les plus consommatrices :
 - Dans ce cas, il faut se référer à la page [Identification des checks les plus consommateurs à l'aide de la commande shinken-scheduler-export-data](#)

Erreurs possibles

La commande est lancée depuis un serveur autre que l'Arbiter

La commande échange avec les Schedulers de la même manière que l'Arbiter. Elle a donc besoin d'avoir les accès réseaux. C'est pour cela que le lancement de la commande n'est autorisé que sur le serveur de l'Arbiter, car c'est le seul à avoir tous les accès vers les démons.

Dans le cas où la commande est lancée depuis un serveur qui n'est pas Arbiter, elle retourne l'erreur suivante :

```
ERROR: command must be launched on the Arbiter server
```

Problème de paramétrage du mot de passe d'accès (daemon__export_data__password dans /etc /shinken/shinken.cfg)

Par défaut, la commande va extraire les données avec les noms des éléments inclus (*désactivé avec la paramètre --anonymous*). Vu que le nom des éléments est une information sensible à protéger, l'accès à ces informations est protégée par un système de mot de passe (*paramètre daemon__export_data__password dans /etc/shinken/shinken.cfg*).

Le mot de passe est vide ou non configuré dans le fichier /etc/shinken/shinken.cfg

Dans le cas où le mot de passe est vide ou non paramétré (*le paramètre daemon__export_data__password dans le fichier /etc/shinken/shinken.cfg*) vous aurez l'erreur suivante :

```
Cannot fetch SCHEDULER      : The "daemon__export_data__password" parameter is missing or void, that is forbidden for not anonymous request
```

Le mot de passe fourni par l'utilisateur est incorrect

Dans le cas où le mot de passe rentré lors du lancement de la commande est incorrect, la commande retourne l'erreur suivante :

```
Cannot fetch SCHEDULER      : The password is not matching "daemon__export_data__password" parameter
```

Problème de paramétrage de la désactivation de l'export des données des démons (scheduler__export_data__enabled parameter dans le .cfg du Scheduler)

Il est possible de désactiver entièrement l'export de donnée sur un Scheduler (*depuis le paramètre `scheduler__export_data__enabled` dans le fichier de configuration du Scheduler*).

Dans le cas où on tente de communiquer avec un Scheduler avec cette configuration, la commande va retourner une erreur :

```
Cannot fetch SCHEDULER      : The Scheduler configuration disabled data export with the "scheduler__export_data__enabled" parameter.
```

Problèmes avec le réseaux ou l'état du Scheduler

Le Scheduler n'est pas accessible (éteint ou problème réseau)

Dans le cas où le Scheduler est éteint ou s'il y a des problèmes réseaux entre le serveur de l'Arbiter et de dernier, l'erreur réseau de connexion sera retournée :

```
Cannot fetch SCHEDULER : ERREUR RESEAU ( connexion refusée, nom d'hôte inconnu, temps écoulé, ... )
```

Le Scheduler n'a pas encore reçu de configuration de l'Arbiter

Si le Scheduler n'a pas encore reçu de configuration de l'Arbiter:

- il n'a aucun éléments supervisés
- et il n'a pas reçu le mot de passe de vérification.

Dans le cas où la commande tente de se connecter au Scheduler avant qu'il ait reçu une communication, l'erreur suivante sera retournée :

```
Cannot fetch SCHEDULER      : The Scheduler does not have any configuration
```

Passer les erreurs et travailler avec des données partielles

Toutes les erreurs concernant les Schedulers sont par défaut bloquantes pour la commande afin d'être sûr d'avoir l'ensemble des données. Si un seul des Schedulers a une erreur, alors la commande sort en erreur et ne génère pas de fichier d'export csv.

Il est cependant possible, via le paramètre `--dont-stop-on-scheduler-error`, de signaler à la commande qu'on accepte les données partielles et que les erreurs sur les Schedulers ne sont que des avertissements et pas des erreurs bloquantes.

Dans le cas par défaut où on ne mets pas le paramètre, la commande indique qu'il est possible de les passer avec le message suivant :

```
There was some errors, cannot continue. Use --dont-stop-on-scheduler-error to allow partial results
```

Si on utilise le paramètre et qu'il y a eu des erreurs qu'on considère comme non bloquantes, alors :

- Les erreurs seront transformées en WARNING au niveau des Schedulers.
- La commande averti tout de même que les données risques d'être partielles avec le message suivant :

WARNING: The results may be partial because some schedulers were in errors: SCHEDULERS