

Compte-Rendu

Auteurs : Diedhiou Ahmed Bachir ,Cisse Mamadou

I.Manipulation des données

a)Présentation des données

Les données se trouvent dans un dossier nommé MESOCENTRE
ce dossier contient à son tour un dossier Nagios_rrd_data contenant les données relevées par l'outil Nagios .

Le dossier Nagios_rrd_data est composé de sous-dossiers contenant les données relatives à chaque hpc et à la climatisation .

Nagios_rrd_data contient trois types de sous-dossiers :

- Un premier dont le nom est de la forme hpc-n* (* = numéro du hpc)
contient les données hpc sous format rrd

- Un deuxième qui a pour nom localhost contient les données rrd sur l'état du chiller & freecooler , la puissance de refroidissement des LCP et la puissance électrique globale

- et enfin hpc-clim-lcp* (= numero du lcp) contenant les données rrd de la température de chaque LCP

▼	MESOCENTRE	--	Dossier
▶	TESTING	--	Dossier
▼	Nagios_rrd_data	--	Dossier
▼	hpc-n315	--	Dossier
	HOST.rrd	768 Ko	Documen
	HOST.xml	15 Ko	XML
▼	localhost	--	
	HOST.rrd.bak	768 Ko	
	HOST.xml.bak	12 Ko	
	Etat_du_chiller_et_du_free_cooler.rrd	768 Ko	
	Etat_du_chiller_et_du_free_cooler.xml	13 Ko	
	PUE.rrd	385 Ko	
	PUE.xml	12 Ko	
	Puissance_de_refroidissement_des_LCPs_en_KW-_LCPs_5_a_8.rrd	1,5 Mo	
	Puissance_de_refroidissement_des_LCPs_en_KW-_LCPs_5_a_8.xml	15 Ko	
	Puissance_de_refroidissement_des_LCPs_en_KW.rrd	1,5 Mo	
	Puissance_de_refroidissement_des_LCPs_en_KW.rrd.bak	1,5 Mo	
	Puissance_de_refroidissement_des_LCPs_en_KW.xml	15 Ko	
	Puissance_de_refroidissement_des_LCPs_en_KW.xml.bak	15 Ko	
	Puissance_electrique_globale_en_KW.rrd	385 Ko	
	Puissance_electrique_globale_en_KW.rrd.bak	385 Ko	
	Puissance_electrique_globale_en_KW.xml	12 Ko	
▼	hpc-clim-lcp2	--	
	Temperatures_du_LCP_2.rrd	2,3 Mo	
	Temperatures_du_LCP_2.xml	16 Ko	

b) rrdtool

RRDTOOLS est un outil de gestion de base de données RRD (Robin-Round Database).

Elle permet de conserver des données dans une base RRD ,tout en incluant un outil permettant de représenter graphiquement des données contenues dans la base .

Notre première approche pour comprendre la structures des données sera d'utiliser RRDTOOLS pour visualiser les graphique de ces données .

Prenons par exemple le fichier rrd du hpc-n101 qui s'appelle _HOST_.rrd

Pour connaître la date de la première mise à jour de la (ou des) première (s) donnée (s)

```
Air-de-Bach:hpc-n101 Bachir$ rrdtool first _HOST_.rrd
1400575140
```

Pour connaître la date de la dernière mise à jour de la (ou des) première (s) donnée (s)

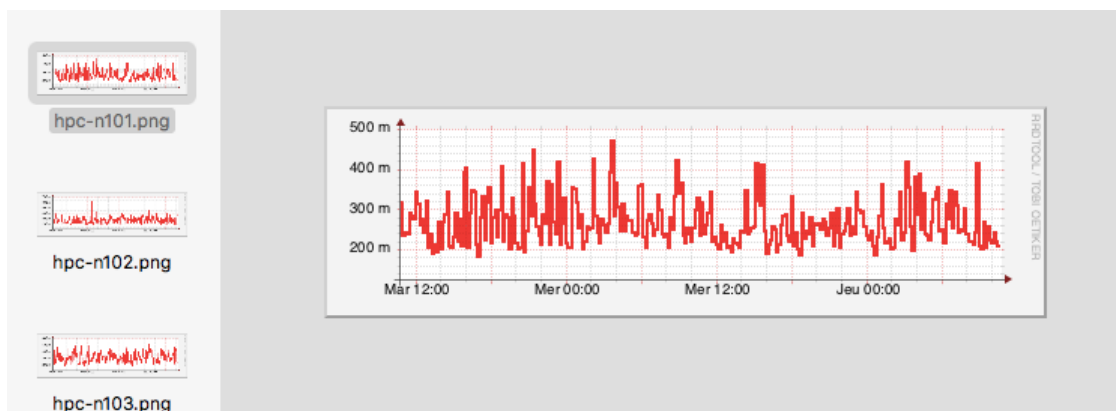
```
Air-de-Bach:hpc-n101 Bachir$ rrdtool last _HOST_.rrd
1400747916
```

Pour visualiser le fichier rrd (on choisit d'afficher les données depuis la première mise à jour j'usqu'à)

```
Air-de-Bach:hpc-n101 Bachir$ rrdtool fetch _HOST_.rrd AVERAGE --start 1400575140 --end 1400575740
1
2

1400575200: 2,0705000000e-01 0,0000000000e+00
1400575260: 2,0230666667e-01 0,0000000000e+00
1400575320: 2,0230666667e-01 0,0000000000e+00
1400575380: 2,0230666667e-01 0,0000000000e+00
1400575440: 2,0230666667e-01 0,0000000000e+00
1400575500: 2,0230666667e-01 0,0000000000e+00
1400575560: 3,3895833333e-01 0,0000000000e+00
1400575620: 3,3895833333e-01 0,0000000000e+00
1400575680: 3,3895833333e-01 0,0000000000e+00
1400575740: 3,3895833333e-01 0,0000000000e+00
1400575800: 3,3895833333e-01 0,0000000000e+00
```

Graphique de quelques rrd des hpc



*Un script bash a été crée pour la génération des graphiques

Etude de l'article « A novel methodology for knowledge discovery through mining associations between building operational data »

Cet article détaille le développement d'une nouvelle méthodologie pour examiner toutes les association/corrélations entre les données opérationnelles (chauffage -ventilation-climatisation) et climatiques (température plein air, humidité), dans le but d'extraire des informations utiles pour la réduction de la consommation d'énergie dans les bâtiments.

Cette méthode est basée sur une méthode de data mining (ARM en anglais) qui est la recherche des règles d'association afin de découvrir des relations intéressantes entre un ou plusieurs paramètres stockés dans la base de données.

Pour comprendre l'ARM, voir: https://en.wikipedia.org/wiki/Association_rule_learning#Definition

Dans cet article le logiciel RapidMiner a été utilisé en même l'algorithme FP-growth pour générer les règles d'association

RapidMiner : <https://rapidminer.com>

FP-growth : <https://fr.scribd.com/document/29371845/Extraction-des-regles-d-association-FP-Growth>

<http://blog.khaledtannir.net/2012/07/lalgorithme-fp-growth-construction-du-fp-tree-23/#.WA9VSqPpNmA>

Méthodologie

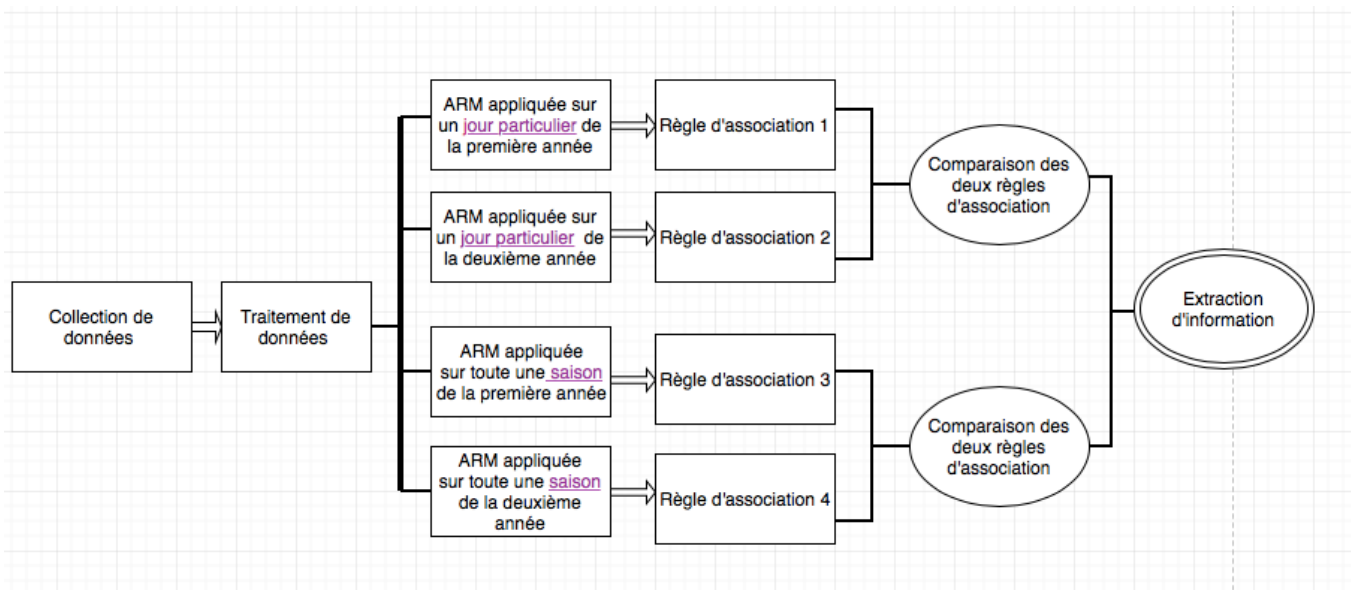


Fig1

Jour particulier = jour le plus chaud, jour le plus froid etc ...

saison = été, hiver, printemps etc

Ce qui est important à savoir concernant la méthode, c'est que lorsqu'une ARM est effectuée sur tous les paramètres des données de la première année, on doit obligatoirement effectuer, lors de la deuxième année, une ARM en utilisant que les paramètres qui sont intervenus dans les règles d'association trouvées avec l'ARM de la première année.

Collection des données

Les données utilisées sont venues d'un pavillon EV situé à Montréal. Elles ont été relevées à partir des différents FHU, RHU et EHU du bâtiment.

Les données de l'année 2007 seront appelées *dataset_1* et celles de 2009 *dataset_2*.

Traitement des données

Classification des données en deux groupes :

-les données sont étiquetées LOW ou HIGH suivant l'intervalle (fixé par les auteurs) auquel elles appartiennent. Par ex : On peut considérer qu'une donnée $\in [0,5]$ est LOW, une autre dans $[5,10]$ est HIGH.

Parfois on peut trouver dans la base de données des valeurs aberrantes (trop grandes ou petites par rapport aux autres valeurs se trouvant dans les données), dans ce cas les auteurs ont établi deux règles pour gérer ces types de données :

Règle 1: les valeurs plus petites que $Q_1 - 1.5 \times (Q_3 - Q_1)$ sont supprimées

Règle 1: les valeurs plus grandes que $Q_3 + 1.5 \times (Q_3 - Q_1)$ sont supprimées

ARM appliquée à un jour particulier de la première année et un jour particulier de la deuxième année

Dans cette étude l'ARM a été d'abord effectuée sur les données du jour le plus froid des années 2007 et 2009. Le *support* a été fixé à 80 %, la *confidence* à 95 %, et le *lift* à un seuil minimal de 1.

Résultat :

L'ARM effectuée sur *dataset_1* a généré 476 règles d'association (**règle d'association 1**, voir Fig1) et 43 paramètres sont mis en jeu.

Ces 43 paramètres ont été utilisés pour effectuer L'ARM sur *dataset_2* et 169 règles d'association (**règle d'association 2**, voir Fig1) ont été générées.

Les règles d'association 1 et 2 sont ensuite comparées, de cette comparaison naîtront trois règles d'association considérées intéressantes.

Ces trois règles leur ont permis de détecter que durant l'hiver la température du FHU4 augmentait d'abord, pour ensuite considérablement diminuer avant alimentation des salles, ce qui représente une perte d'énergie. Les auteurs ont pu remarquer que cette baisse de température était due à la température (2 °C) de l'eau que l'humidificateur propageait. Après avoir visité le site leurs remarques ont été confirmées. À partir de ces remarques une nouvelle stratégie a été adoptée pour résoudre ce problème.

ARM appliquée sur toute une saison de la première année et une saison la deuxième année

Ici l'ARM est effectuée sur les données de l'hiver 2007 et 2009 .

Les données de 2007 sont toujours appelées *dataset_1* et celles de 2009 *dataset_2*

support =50 %

confidence=80 %

lift =1

L'ARM effectuée sur *dataset_1* a généré 461 règles d'association (**règle d'association 3** , voir Fig1) et 32 paramètres sont mis en jeu .

Ces 32 paramètres ont été utilisé pour effectuer L'ARM sur *dataset_2* et 262 règles d'association (**règle d'association 2** , voir Fig1) on été générées .

Les règles d'association 2 et 3 sont ensuite comparées , de cette comparaison naitra trois règles d'association qui seront classifiées en 3 catégories de règles

La première catégorie de règle a permis de voir que le débit d'air du FHU 1 et celui du FHU2 ont tendance à varier de la même manière en 2007 ,puis en 2009 , et de voir les valeurs de débits de ces deux FHU sont sensiblement proches durant l'hiver de ces deux années.

En conclusion cette catégorie leur permettra d'inférer qu'il ya une défaillance des FHU lorsque les valeurs de débits observés présenteront une grande différence .

Les deux catégories de règles restantes leur ont aussi permis de tirer des informations intéressantes.

Conclusion:

Le but de ce premier travail a été de maitriser rrdtool ,d'écrire un script pour la visualisation graphique des données et de lire l'article pour essayer comprendre la méthode utilisée pour l'analyse des données .

Notre première approche a été de créer un script bash pour automatiser le tracé graphique des différentes données qui sont à notre disposition .Un script équivalent sera prochainement écrit en python.

La méthode présentée dans l'article est vraiment intéressante quand on veut déterminer les associations et corrélations entre les différents paramètres d'une base données.

Nous essayerons de nous en inspirer pour la mise en œuvre de notre projet .Ceci étant ,nous espérons que la lecture d'autres articles nous aidera à avoir des outils importants pour une analyse efficace des données que nous possédons .