

Proposition de roadmap

Rapport Altran-smart energy

Idée générale : le projet est comme une mini-mission de consulting en data science, où, avant de travailler sur les données, il faut bien comprendre le contexte général (état de l'art, autres approches existantes) et la situation physique (le data center dont on va étudier les données)...

Proposition de roadmap pour le rapport :

I. Contexte et objectif

1.1 Efficacité énergétique pour les data centers

1.1.1 Motivation

Chercher quelques stats sur le volume de données stockées + leur augmentation (perspectives)

Consommation énergétique actuelle? A venir? --> par ex. trouver des stats du style

"la conso énergétique actuelle correspond à la production de X centrales nucléaires pdt Y mois" ou

"les data centers de google consomment actuellement autant que Z foyers"...

Il y a qq pistes ici:

http://www.lesechos.fr/19/11/2013/LesEchos/21566-050-ECH_comment-rendre-les---data-centers--plus-sobres.htm

1.1.2 Solutions actuelles

Donner un aperçu par quelques ex. de ce qui est mis en place pour réduire la conso des data centers ou réutiliser leur énergie.

Par ex. Celeste (green data centers), google environnement, il existe des start-ups qui proposent de mettre des data centers en sous-sol d'immeuble. Cf les refs sur Asana.

Un projet intéressant à citer: une piscine chauffée par un data center

<http://www.20minutes.fr/paris/1800039-20160304-paris-quand-data-centers-chaufferont-ville>

1.2 Cas d'étude

Présenter le data center. Ref technique ici:

<https://alsacalcul.unistra.fr/nos-services/presentation-du-cluster/>

CR de la visite: décrire les installations, faire le schéma de la situation physique → recycler rapport #2.

1.3. Bonus: quelques préconisations physiques

Quelques pistes sur lesquelles un data center peut gagner:

Cf les tickets sur Asana (optimiser % glycol dans les circuits, débit d'air, température de consigne vs abaques constructeurs etc)

II. Les Données

Maintenant qu'on comprend mieux le système qu'on va étudier, on peut se pencher sur les données

2.1 Les données dont on dispose et leur stockage

2.1.1 Format des données: rrd

cf rapport #1: round robin data base: principe de fonctionnement?

Expliquer brièvement le mode d'intégration et stockage: Nagios, rrd tools

2.1.2 Description

Décrire les jeux de données + mettre en lien avec le schéma physique du data center (tableau des points de mesure etc) → rapports #1 et #2

2.2 Traitement des données

Ici, des résultats: scripts pour préparer les données (formatage en csv, script pour ploter en rrd tools...)

Montrer quelques plots, expliquer le système d'horodatage... (cf rapport #1)

III. Règles d'associations

Idée générale: visualiser des liens causes conséquences (si ils existent !)

3.1 Motivation: description de l'article

Décrire la méthode utilisée dans l'article de Yu et al.

3.2 Méthodes utilisées

Décrire le ou les algos testés (+ éventuellement librairies si R ou Python...)

3.3 Résultats:

Association map à partir des données ?

IV - Modélisation

A partir des data de rrd tools, on peut tracer des plots par ex. température en fonction du temps. Créer un script (R, Python...) qui pour chaque plot permet d'extraire un histogramme (Fig 1)

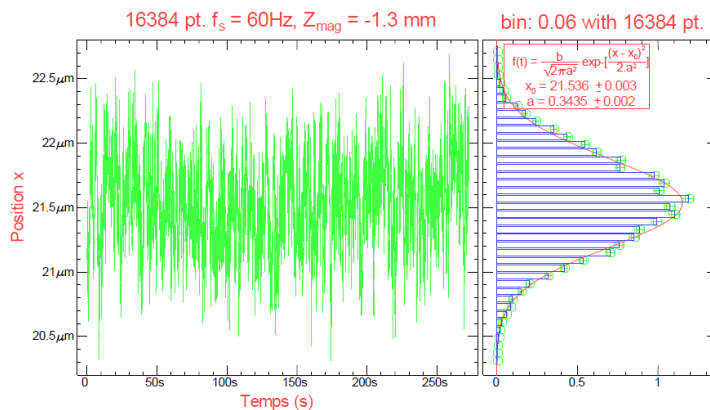


FIGURE 1 – A gauche, signal expérimental : position $x(t)$ d'une bille soumise à une perturbation brownienne au cours du temps. Ce signal contient 16384 points, il est échantillonné à 60Hz. A droite, histogramme de ce signal indiquant que la distribution des positions est gaussienne.

Fig 1. Données, histogramme et sa distribution.

Peut-on fitter des histogrammes? → nature de la distribution?
Commentaires sur le résultat?