

# Gestion des logs au CCIN2P3 et passage à l'échelle : le ML à la rescousse ?

*jeudi 29 mars 2018 15:15 (25 minutes)*

Près de 100 millions de “logs” et 1 milliard de métriques sont collectées par jour dans les deux datacentres du CCIN2P3. Ils sont traités via une plateforme implémentant le modèle dit “lambda” : les événements traversent deux “pipeline” en parallèle. La première de faible latence permet un traitement synchrone, presque temps-réel. La deuxième permet un traitement “batch” asynchrone sur toute ou partie des événements passés.

Les deux “pipeline” permettent de notifier les gestionnaires de service du bon ou mauvais fonctionnement de leurs services grâce à des algorithmes et des règles prédéfinis, at modifiables à souhait.

Le problème de cette approche est qu'elle ne passe pas à l'échelle. En effet, la quantité de “logs” ne fait qu'augmenter, et le temps humain nécessaire à maintenir le jeu de règles et d'algorithmes qui permettent de détecter les problèmes également.

Le besoin de trouver une autre stratégie se fait sentir, et les techniques de ML voire de DL semblent prometteuses pour assister par exemple en utilisant les techniques de “outlier detection”.

**Author:** WERNLI, Fabien (Sysadmin)

**Orateur:** WERNLI, Fabien (Sysadmin)