

Construire et maintenir une plateforme de calcul R avec JupyterHub

David Gohel 1*

Résumé

Utiliser R *en production* dans une organisation nécessite l'administration d'un écosystème R. La mise à disposition au sein d'un système d'information existant de l'outil R et plus généralement d'outils open source pour la data-science est une tâche complexe. Il faut maîtriser les versions mises à dispositions, permettre l'utilisation et l'accès aux données dans un cadre sécurisé, permettre les mises à jour fréquentes de l'outil et des packages utilisés. Il existe une solution pour aider à cette administration, le projet open source JupyterHub qui va être présenté à travers un cas réel d'utilisation.

Mots-clefs : R – production – serveur – docker – kubernetes

Développement

JupyterHub nous aide à créer et administrer notre plateforme R en offrant aux utilisateurs des environnements prêts à l'emploi et aux administrateurs les outils nécessaires pour l'intégration et la maintenance de l'application.

L'outil permet d'embarquer les logiciels R, python, Julia, bash, etc. On peut y intégrer les notebooks mais aussi RStudio et bien d'autres outils complémentaires. L'interface utilisateur est une interface web et n'exige de l'utilisateur qu'un navigateur WEB.

1. Je présenterai les caractéristiques de JupyterHub, c'est à dire, sa flexibilité, ses capacités de personnalisation, son évolutivité et sa portabilité.
2. La démonstration d'un cas réel d'utilisation de la solution par "CEVA Santé Animale" sera réalisée. Je montrerai la plateforme avec R, RStudio et les notebooks et discuterai de l'interface JupyterLab et de la richesse des outils qui peuvent s'y intégrer.
3. Le déploiement de la solution sera enfin discuté. Je mettrai l'accent sur les outils d'intégration continue impliqués dans le processus de déploiement, comme la gestion des images docker et l'automatisation de la construction des environnements utilisateurs.

*ArData, david.gohel@ardata.fr