



HAL
open science

RunMyCode

Yvan Stroppa

► **To cite this version:**

Yvan Stroppa. RunMyCode. journées scientifiques mésocentres et France Grilles 2012, Oct 2012, Paris, France. hal-00766076

HAL Id: hal-00766076

<https://hal.science/hal-00766076>

Submitted on 17 Dec 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PLATE-FORME RunMyCode

G. COLLETAZ, C. HURLIN, C. PERIGNON, V. STODDEN, Y. STROPPA,
christophe.hurlin@univ-orleans.fr, LEO, Université d'Orléans
gilbert.colletaz@univ-orleans.fr, LEO, Université d'Orléans
perignon@hec.fr, GREGHEC, HEC(Paris)
victoria@stodden.net, Columbia New-York
Yvan.stroppa@univ-orleans.fr, LEO, CNRS, INSHS

Overview:

RunMyCode is a web service allowing people to run computer codes associated with a scientific publication (articles and working papers) using their own data and parameter values. The service only requires a web browser as all calculations are done on a dedicated "cloud computer". Once the results are ready, they are automatically displayed to the user. RunMyCode is a non-for-profit scientific website founded by academics and funded by several national research agencies and universities. The goal of the website is to make academic research easier to use and to replicate. RunMyCode relies on the concept of a *paper's companion website*.

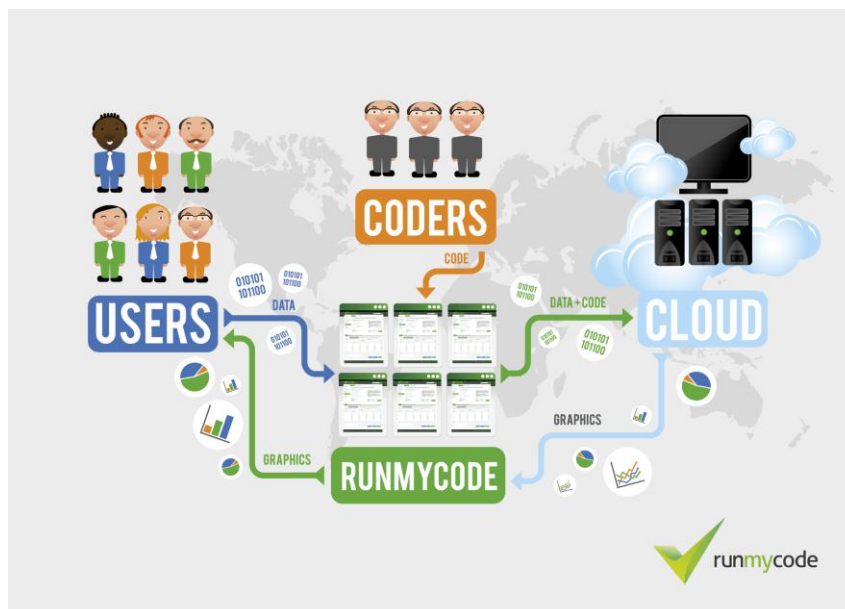


Figure 1: représentation globale de l'architecture RunMyCode

Enjeux scientifiques, besoin en calcul, stockage et visualisation :

Les objectifs de la plate-forme sont :

- permettre aux chercheurs français et étrangers de **diffuser** de façon rapide à l'échelle internationale les résultats de leurs recherches à partir d'un site centralisant des ressources informatiques de calcul et de traitement, disponibles en ligne. Ce service devrait notamment permettre d'augmenter le potentiel de citations de certains articles scientifiques.
- Offrir à une très large communauté d'utilisateurs, qui dépasse largement la seule communauté académique, la possibilité d'utiliser, en ligne et de façon **interactive**, des méthodes de calcul et de traitement scientifiques parmi les plus récentes pour leurs propres données et avec leurs propres choix de paramètres. Cette utilisation n'était jusque là pas possible pour l'ensemble des utilisateurs qui ne possédaient pas les compétences informatiques pour implémenter ces méthodes sur des progiciels dédiés.
- Fournir à la communauté académique (chercheurs, éditeurs, arbitres, etc.) un moyen de pouvoir **répliquer** les résultats scientifiques et d'en démontrer ainsi l'exactitude, la fiabilité et la portée générale.

L'architecture de RunMyCode est de type MOM (Middleware Oriented Messages), elle est organisée en deux parties : une partie frontale et une partie Backend. La communication entre les deux parties se fait principalement par envoi de messages et de façon asynchrone. Chaque message envoyé par l'application CompanionsSite à la partie Backend contient tous les paramètres nécessaires pour le calcul. La partie Backend contient l'ensemble des solveurs et le contexte d'exécution (Pré et post-traitement sous Perl) associé à chaque site Compagnon. Elle reçoit le message et le traite jusqu'à la récupération par le frontal du fichier résultat.

Les besoins de calculs (en temps CPU, en espace de stockage et en quantité mémoire) sont variables en fonction des sites sur lesquels les internautes interagissent et sont conditionnés par les caractéristiques des files d'attente définies au sein des centres de calculs. Le module supplémentaire nommé DTM (Distributed Task Manager) permet de stocker l'ensemble des demandes dans une file d'attente spécifique et de les consommer à **flux maîtrisé** par l'activation d'agents résidents sur l'infrastructure de calcul. Ce système permet une consommation limitée des ressources. De plus, un mécanisme de cache (sous IRODS) permet avant toute exécution de vérifier si le résultat demandé a déjà été calculé. Dans ce cas et si l'auteur du site l'a autorisé, le résultat est restitué à l'internaute.

La visualisation des résultats de calculs est actuellement obtenue sous forme de fichier pdf, une version est à l'étude pour visualiser directement dans le navigateur les résultats en web 2.0 (plus grande interaction).

RunMyCode apporte en plus du concept site compagnon à une publication ou un working paper le concept de **site thématique** qui permet aux internautes (à la communauté scientifique) de disposer de moyen d'exécution sur plusieurs moteurs de calculs (plusieurs sites compagnons) avec un concept de Map-Reduce avancé pour s'appuyer pleinement sur une architecture distribuée.

Illustration d'une demande de soumission simple

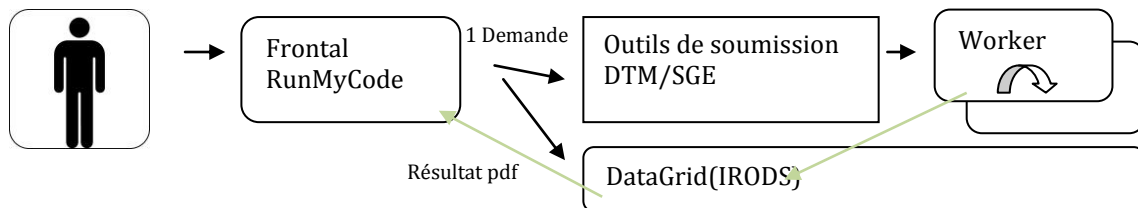
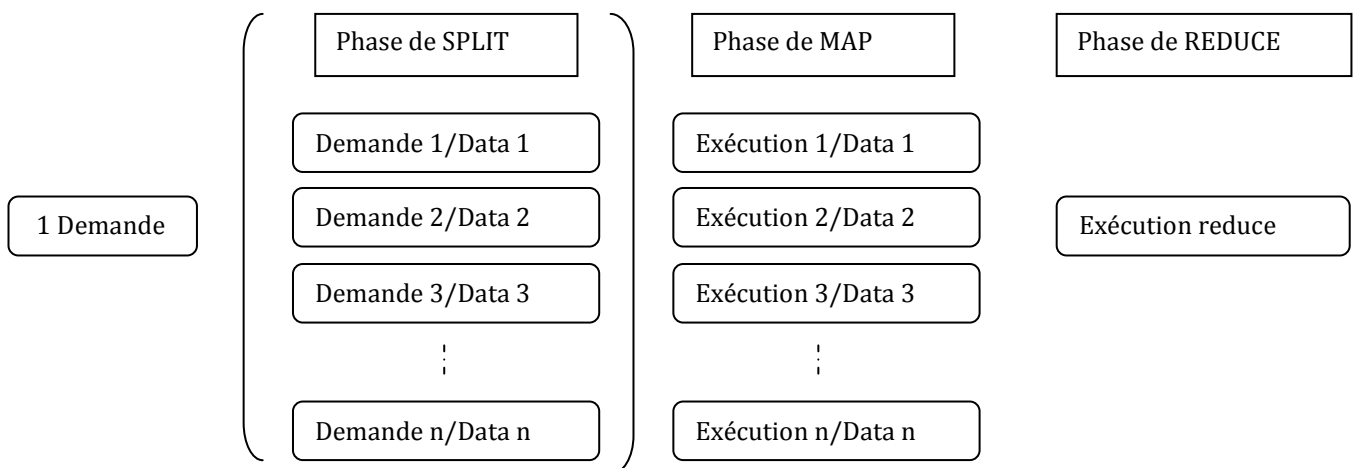


Illustration d'une demande de soumission dans le cadre d'un site thématique/RO (Recherche opérationnelle)



La phase de split a lieu dans le cadre de site RO, car les paramètres fournis par l'utilisateur sont des expressions (exemple sur Matlab $0:1:10$ → ce qui indique que le paramètre doit varier par pas de 1 de 0 à 10)

En résumé : la plate-forme peut traiter des demandes

- Un jeu d'inputs avec un solveur
- Un jeu d'inputs avec plusieurs solveurs
- Plusieurs jeux d'inputs avec un solveur
- Plusieurs jeux d'inputs avec plusieurs solveurs

Développements, utilisation des infrastructures :

Situation actuelle :

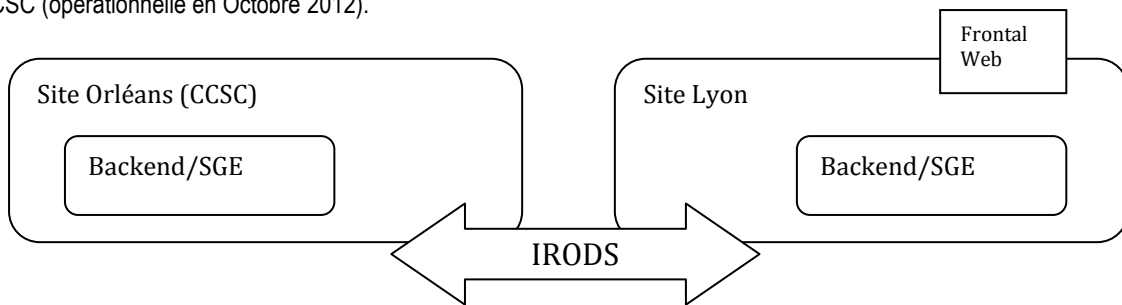
Depuis fin 2010, le développement de la plate-forme a été réalisé au sein laboratoire LEO UMR7322 par l'équipe RMC'LABS en collaboration avec l'équipe du CCIN2P3 à Lyon et du TGE-ADONIS. La plate-forme à ce jour est opérationnelle et permet aux chercheurs de constituer leurs sites Companions et à l'équipe RunMyCode de les intégrer. Actuellement plus de 60 Sites Companions sont disponibles et couvre deux champs disciplinaires l'ECO-Gestion et les Mathématiques (traitement d'images).

Développement :

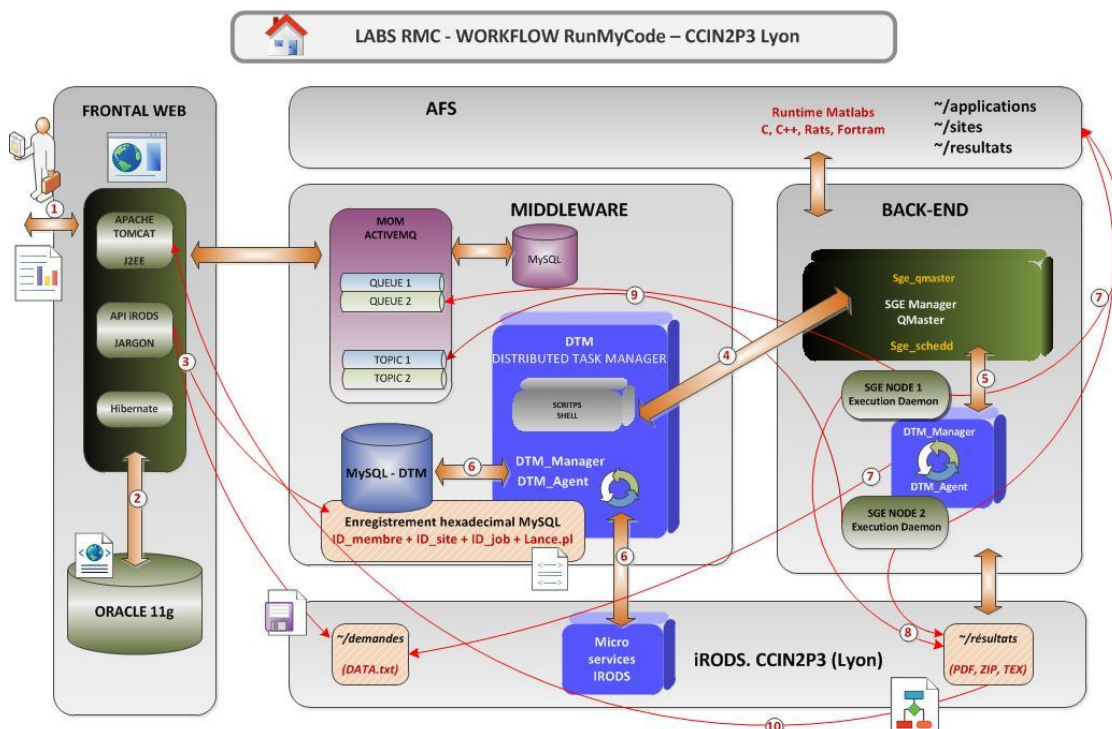
RMC'LABS en collaboration avec les autres membres du projet, définit et développe les différentes applications aussi bien au niveau du frontal (partie WEB) que de la partie backend (logistique, traitement et exécution des tâches). Les applications sont développées sous l'environnement J2EE et s'exécutent sous TOMCAT.

La plate-forme est abritée sous l'infrastructure mutualisée du TGE-Adonis (INSHS) Elle s'appuie sur un processus de gestion de tâches DTM (Distributed Task Manager) qui permet aux applications Tomcat de soumettre des demandes d'exécution sur les serveurs de calculs

Cette mise en place et ces choix techniques permettent d'envisager un couplage de plusieurs centres de calculs afin d'étendre l'offre de calculs de RunMyCode. Un couplage est en phase de prototypage avec le centre de calculs d'Orléans CCSC (opérationnelle en Octobre 2012).



Des développements sont élaborés conjointement avec plusieurs structures telles que le LI de Tours (Laboratoire Informatique) et l'EFREI (Ecole d'ingénieurs Villejuif). Ces partenaires contribuent dans l'exploration de solution permettant de répondre aux nombreuses problématiques que soulève ce projet.



Outils, le cas échéant complémentarité des ressources, difficultés rencontrées :

Un ensemble d'outils est mis en place afin de permettre une grande robustesse de la plate-forme RunMyCode. Parmi ces outils on peut citer ActiveMQ et MongoDB qui devraient permettre la mise en place d'une logistique calcul beaucoup plus fine et l'alimentation d'un système de supervision propre à la plate-forme pour garantir une haute disponibilité. De plus, ils permettront de mettre en place des mécanismes de répliquions entre différents sites calculs.

Plusieurs difficultés sont à surmonter :

- une automatisation du processus d'encapsulation des codes,
- une performance à la hauteur des attentes des internautes et des auteurs,
- Les calculs distribués sur les centres de calculs pour répondre aux sites thématiques et RO,
- L'ordonnancement des demandes afin de desservir aussi bien un internaute qui demande l'exécution d'un calcul rapide que celui qui demande un ensemble de calculs importants. D'où la collaboration avec le Laboratoire d'Informatique de Tours pour les problématiques liées à l'ordonnancement.

Résultats scientifiques :

Les résultats scientifiques espérés sont une accélération de la diffusion et une valorisation de la production scientifique. Une meilleure utilisation des méthodes et des programmes développés, la possibilité de répliquer les résultats de calculs et de comparer différentes approches.

Permettre aux chercheurs d'avoir des feedback direct ou par reporting sur leurs différentes réalisations ; vérifier la robustesse d'un algorithme, d'une méthode face à de nombreuses sollicitations, permettre aux différents acteurs d'échanger dans des espaces en dehors ou au sein de la plate-forme (espace collaboratif de différents types – blog, forum, wiki, développement de type git ...).

L'appui de certains éditeurs pourrait favoriser l'émergence de cette solution dans la sphère académique et ainsi améliorer les processus de validation, de diffusion et de valorisation de la recherche.

Permettre de créer une dynamique autour de cette plate-forme entre les différents types utilisateurs (Les auteurs de site et les utilisateurs de ces sites).

Permettre le partage de données, de résultats et ainsi la possibilité de constituer des banques de données (d'images, de fichiers texte ...). Eléments de références permettant de comparer les effets et améliorations possibles de traitements nouveaux.

Mise en place d'une communauté et ainsi favoriser l'émergence de collaboration autour de problématique commune.

Perspectives :

Les perspectives de déploiement : au sein de la discipline Economie-Gestion à hauteur de 100 sites compagnons d'ici fin d'année 2012, ce nombre devrait avec l'apport d'une collaboration avec les éditeurs ELSEVIER, SSRN connaître un accroissement considérable pour les années suivantes. On compte dans cette seule discipline entre 800 et 1000 sites compagnons par an si on se base sur l'ensemble des revues SHS. Le tout pouvant être encore augmenté avec l'arrivée d'ici 2013 de l'accompagnement de conférences ou colloques qui devrait permettre de voir l'émergence de ce concept dans cette sphère scientifique.

Le déploiement dans les autres disciplines est prévu pour Septembre 2012 dans le champ disciplinaires des Mathématiques et dans le champ disciplinaire de la Bio-Statistiques prévu février 2013. Cette action devrait amener dans sa première année environ 200 sites compagnons supplémentaires.

Les autres champs disciplinaires ont déjà été rapprochés et sont en cours de spécifications. Pour une intégration prévue 2013.

Développement communautaire : des perspectives de mise en place autour de cette plate-forme d'un système de collaboration devrait pouvoir accroître l'intérêt des échanges autour de ces productions numériques. Dans cette phase, le

partage à l'échelle mondiale devrait amener un nouveau moyen de communication scientifique.

Développement et couplage géographique :

Le couplage multi-sites doit permettre de supporter cette demande croissante dans ce milieu par la mise en place de ce concept. On prévoit fin d'année 2012 un couplage effectif entre le frontal de TGE-ADONIS et le centre de calculs d'Orléans (CCSC).

Le développement d'un réseau social autour de ce concept permettant et facilitant l'émergence d'échange et de collaboration entre les différents acteurs est en cours d'étude devrait voir le jour courant 2013.

Références :

[Ince, Hatton and Graham-Cumming, "The Case for Open Computer Programs", Nature, February 2012, 482, 485-488](#)

[Stodden V. and Arbesman S., "Scientists, Share Secrets or Lose Funding", Bloomberg View, Jan 10, 2012](#)

[Morin et al, "Shining Light into Black Boxes", Science, April 2012](#)