



HAL
open science

FG-Cloud : un service de cloud computing fédéré pour le calcul scientifique

Cécile Cavet, Vincent Legoll, Jérôme Pansanel, Sorina Camarasu-Pop, Adrien Ramparison, Geneviève Romier, François Thiebolt

► **To cite this version:**

Cécile Cavet, Vincent Legoll, Jérôme Pansanel, Sorina Camarasu-Pop, Adrien Ramparison, et al.. FG-Cloud : un service de cloud computing fédéré pour le calcul scientifique. Journées Réseaux de l'Enseignement et de la Recherche.(JRES) 2017, Nov 2017, Nantes, France. hal-02019194

HAL Id: hal-02019194

<https://hal.science/hal-02019194>

Submitted on 14 Feb 2019

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.

FG-Cloud : un service de cloud computing fédéré pour le calcul scientifique

Cécile Cavet

Université Paris Diderot-Paris 7
Laboratoire APC
Bâtiment Condorcet
Case 7020
75205 Paris Cedex 13

Vincent Legoll

OpenIO
2 bis, avenue Antoine Pinay
59510 Hem

Jérôme Pansanel

Institut Pluridisciplinaire Hubert Curien
23, rue du Loess – BP28
67037 Strasbourg Cedex 2

Sorina Pop

CREATIS
Bâtiment Blaise Pascal
7, avenue Jean Capelle
69621 Villeurbanne Cedex

Adrien Ramparison

Laboratoire de l'Accélérateur Linéaire
Centre Scientifique d'Orsay
Bâtiment 200 – BP34
91898 Orsay Cedex

Geneviève Romier

Centre de Calcul de l'IN2P3
21, avenue Pierre de Coubertin CS70202
69627 Villeurbanne Cedex

François Thiebolt

Institut de Recherche en Informatique de Toulouse
118, route de Narbonne
31062 Toulouse Cedex 9

Résumé

L'infrastructure de recherche France Grilles, initialement focalisée sur les grilles de calcul et le stockage distribué, propose depuis 2014 une offre de services basée sur le « Cloud Computing » pour le calcul scientifique.

Il s'agit, au travers de l'intégration des dernières avancées technologiques, de proposer un service répondant aux évolutions des besoins de traitement des données scientifiques, notamment en termes d'élasticité, de flexibilité et de simplicité d'accès. Ainsi, afin de proposer un service répondant aux mieux à ces besoins, France Grilles a développé une offre basée sur une fédération de Cloud Openstack, à laquelle est adossé un ensemble de services pour le traitement des données scientifiques.

Le développement de cette fédération, réalisée grâce à la mise en commun de ressources (calcul, stockage, réseau et humaines) géographiquement distribuées en France et fournies par plusieurs sites, est pilotée par un groupe de coordination composé d'administrateurs système, d'utilisateurs experts et de développeurs. Plusieurs développements ont également été réalisés pour pouvoir atteindre le niveau de production (supervision, authentification, distribution des images, ...).

Côté utilisateur, le service propose un ensemble de fonctionnalités en adéquation avec les besoins actuels des chercheurs, des enseignants-chercheurs et des ingénieurs (formations, serveur pour calcul scientifique, Docker as a Service, Spark, Jupyterhub, ...). L'utilisation de l'infrastructure par les scientifiques a déjà permis la publication de plusieurs articles. Enfin, certains sites étant membres du Cloud fédéré EGI, un accompagnement est également proposé aux utilisateurs souhaitant utiliser de ces ressources complémentaires.

Mots-clefs

Cloud, fédération, OpenStack, calcul scientifique, partage d'expertise, France Grilles, EGI, calcul distribué, stockage

1 Introduction

France Grilles,¹ groupement d'intérêt scientifique (GIS) créé en 2010, a pour missions d'établir, d'opérer et de fédérer des moyens de calcul et de stockage géographiquement distribués en France. En 2012, son périmètre d'action a été étendu aux ressources de type « *Cloud Computing* ».² Depuis lors, différents projets ont été menés, permettant d'aboutir à une offre de service pour le calcul scientifique sur des ressources *Cloud* en 2014. Elle se dénomme FG-Cloud.³

La mise en place de cette offre a nécessité une stratégie claire déployée par un groupe de pilotage, composé d'administrateurs système, de développeurs et d'utilisateurs. Elle a permis de fédérer plusieurs infrastructures sans avoir à imposer de contraintes fortes, de créer un réseau d'experts dans le domaine du *Cloud* et de mettre en place des outils pour faciliter l'exploitation du service par les utilisateurs.

Ainsi, des chercheurs issus de plusieurs domaines scientifiques ont déjà pu utiliser ce service afin de mettre en place leur propre infrastructure numérique de recherche dans le *Cloud*.

2 FG-Cloud

2.1 Stratégie

France Grilles vise au développement et à l'opération d'une infrastructure *Cloud* destinée à répondre aux besoins des applications scientifiques. L'accès aux ressources doit être ouvert à toutes les communautés scientifiques et faciliter les approches pluridisciplinaires. Cette fédération de *Cloud* se positionne comme complémentaire des autres acteurs du domaine, qu'ils soient publics ou privés. Elle doit s'intégrer dans une infrastructure internationale de manière à servir les communautés dont le cadre dépasse celui de la France. France Grilles tient à ce titre le rôle de représentant pour la France au sein de ces infrastructures ou projets.

À partir de cette stratégie, les objectifs suivants ont été fixés :

- développer une communauté d'experts des technologies *Cloud* ;
- valoriser leur expérience par le biais de publications et d'interventions lors de meetings et conférences ;
- mutualiser les expériences de chacun afin de faciliter l'adoption des technologies par de nouveaux sites et utilisateurs en diffusant la connaissance et les compétences au travers de formations et documentation publique ;
- créer une infrastructure pouvant être utilisée par la communauté scientifique pour effectuer des travaux de recherche ;
- être le partenaire de choix entre les communautés Françaises et internationales pour toutes les problématiques liées à l'utilisation de Clouds dans des contextes de recherche scientifique ;
- le support aux utilisateurs et l'accompagnement dans la prise en main des technologies *Cloud*.

2.2 Pilotage

Le projet est piloté par un groupe composé d'administrateurs système, de développeurs et d'utilisateurs. Ce groupe se réunit deux fois par mois par visio-conférence, au cours desquelles le point est fait sur l'évolution des sites, le suivi des utilisateurs, les développements en cours ainsi que la programmation des formations.

1. <http://www.france-grilles.fr/>

2. https://fr.wikipedia.org/wiki/Cloud_computing

3. <http://www.france-grilles.fr/catalogue-de-services/fg-cloud/>

Ces réunions permettent de suivre l'évolution de l'infrastructure, d'échanger autour des problèmes d'administration des services et des opérations, et de maintenir l'adéquation du service au besoin des utilisateurs. L'accès à ces réunions est ouvert à toute personne intéressée par l'administration d'une infrastructure *Cloud*, le développement d'outils reposant sur ce type de technologie ou l'utilisation de cette infrastructure dans le cadre d'une activité scientifique.

Des outils collaboratifs ont été mis en place pour mutualiser au mieux expériences et bonnes pratiques (wiki, listes de discussion, hébergement de code).

3 Le service FG-Cloud

3.1 Infrastructure du Cloud fédéré

Le service repose sur une fédération d'infrastructures *Cloud* de type Infrastructure as a Service⁴ (IaaS) hébergées et administrées par une dizaine de sites. Ils permettent aux utilisateurs d'accéder globalement à un ensemble de 6300 cœurs, 20 To de mémoire et 1200 To de stockage. Des ressources sont régulièrement ajoutées pour développer l'infrastructure et répondre aux besoins croissants des utilisateurs.

La solution *OpenStack*,⁵ ensemble de logiciels permettant la mise en place d'une infrastructure *Cloud*, est utilisée dans chacun des sites. L'utilisation de cette suite logicielle, leader dans son domaine, permet de garantir des évolutions fonctionnelles régulières et une intégration facilitée dans les outils des chercheurs. Elle permet également d'accéder de manière uniforme aux ressources.

Les sites participants sont détaillés sur la figure 1.

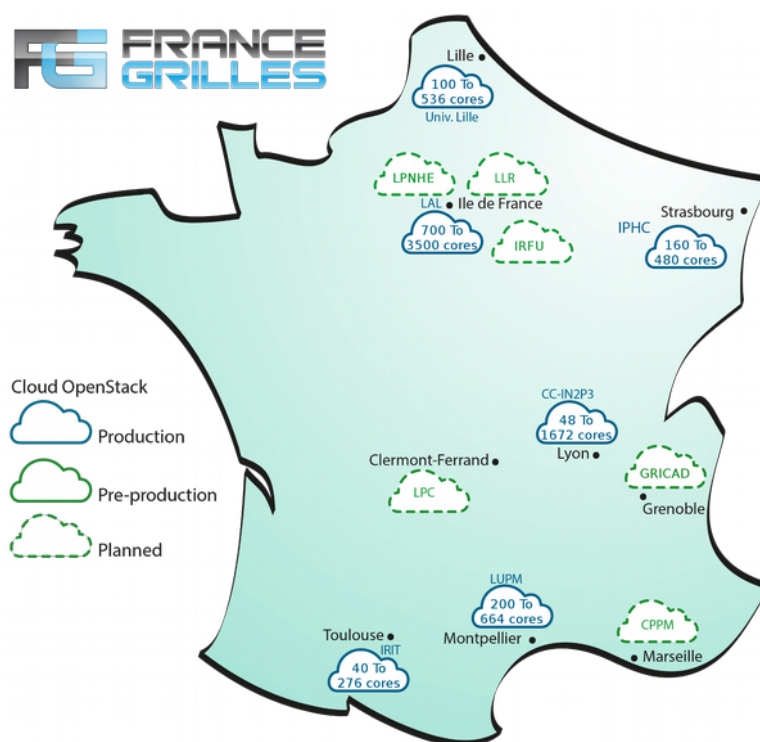


Figure 1 - Carte de France des sites participant à l'infrastructure

4. https://fr.wikipedia.org/wiki/Infrastructure_as_a_service

5. <https://fr.wikipedia.org/wiki/OpenStack>

Ils sont classés en trois catégories :

- en cours de déploiement, regroupant les sites qui sont en cours d'installation ;
- pré-production, ensemble des sites fonctionnels, mais qui n'ont pas encore été intégrés dans la fédération ;
- production, pour les sites fonctionnels au sein de la fédération.

3.2 Extension de l'infrastructure

Deux moyens permettent l'agrandissement de l'infrastructure :

- ajout de nouvelles ressources dans un site déjà existant. Cette solution permet, par exemple, de mutualiser l'administration des serveurs et des infrastructures d'hébergement ;
- intégration d'un nouveau site de cloud computing dans l'infrastructure fédérée.

L'extension de l'infrastructure est discutée lors des réunions du comité de pilotage et un accompagnement particulier est proposé lorsqu'un nouveau site souhaite rejoindre la fédération. Les sites sont libres de leurs choix technologiques, les outils permettant d'intégrer la fédération étant très souples.

3.3 Développements réalisés

Afin de faciliter l'exploitation de l'infrastructure, des outils ont été développés et déployés :

- outil de surveillance fonctionnel basé sur Nagios ;
- comptabilité des ressources utilisées par les utilisateurs France Grilles ;
- distribution automatisée des images ;
- authentification centralisée des utilisateurs ;
- amélioration de la traçabilité des actions pour faire face aux alertes de sécurité.

3.4 Services pour les utilisateurs

Afin de faciliter l'utilisation de FG-Cloud par les utilisateurs, l'offre comporte plusieurs services. Cette liste n'est pas figée et peut être étendue en fonction des nouvelles demandes.

3.4.1 Formation et documentation

Des formations sont organisées à la demande pour les utilisateurs de FG-Cloud. Elles sont de deux types :

- courte (une demi journée) pour se lancer rapidement dans l'utilisation du service ;
- longue (trois jours) pour découvrir en profondeur l'utilisation du *Cloud* et tirer partie du service FG-Cloud dans ses applications scientifiques.

Afin d'accompagner les utilisateurs, une [documentation complète](#) est disponible et le réseau d'expert est volontaire pour les aider à migrer leur application dans le *Cloud*.

3.4.2 Simplification de l'accès au *Cloud*

Par défaut, accéder à un *Cloud* de type IaaS nécessite l'utilisation d'un programme en ligne de commande et l'installation du client OpenStack. Ce type d'interface utilisateur peut être un frein à l'adoption de cette technologie et plusieurs acteurs ont développés leurs propres interfaces pour la gestion des machines virtuelles et des ressources associées. [Horizon](#)⁶ est la plus connue d'entre elles, s'agissant d'un module officiel OpenStack.

6. <https://www.openstack.org/software/releases/ocata/components/horizon>

Dans le cadre de FG-Cloud, trois solutions ont été mises en place :

- des tableaux de bord⁷ OpenStack (horizon) pour accéder spécifiquement à l'un des sites de l'infrastructure ;
- SlipStream⁸ pour le déploiement et l'orchestration de machines virtuelles sur FG-Cloud ;
- DIRAC⁹ pour le lancement de calculs scientifiques sur FG-Cloud.

3.4.3 Accès aux données

Les données analysées ou produites dans le *Cloud* peuvent être stockées sur l'infrastructure FG-iRODS.¹⁰ Elles sont alors accessibles depuis d'autres infrastructures de calcul (grille de calcul, centre de calcul, IDRIS) ou depuis n'importe quel poste client.

3.4.4 Accès au *Cloud* fédéré EGI¹¹

Si les utilisateurs ont des besoins dépassant les ressources disponibles dans l'infrastructure France Grilles, ou s'ils font partie d'une collaboration internationale, ils pourront être accompagnés pour la mise en production de leurs calculs scientifiques sur l'infrastructure du *Cloud* fédéré EGI.¹²

3.5 Lien avec les autres infrastructures *Cloud*

Le groupe FG-Cloud travaille régulièrement avec d'autres projets d'infrastructures *Cloud*, que ce soit au niveau national ou international.

Au niveau français, les sites de l'Université de Lille¹³ et de l'Institut Pluridisciplinaire Hubert Curien¹⁴ (IPHC) sont également membres du *Cloud* de l'Institut Français de Bioinformatique¹⁵ (IFB). Des administrateurs de l'IFB ont pu suivre des formations organisées par France Grilles et inversement.

Au niveau européen, France Grilles suit les activités du *Cloud* fédéré EGI. Le site de l'IPHC est également membre du *Cloud* fédéré EGI. Des développements communs sont réalisés (distribution des images de machines virtuelles, sécurité) et des formations sont organisées pour aider les utilisateurs de FG-Cloud à utiliser l'infrastructure européenne.

4 Utilisation du service

4.1 Accès à FG-Cloud

Le service FG-Cloud, tout comme les autres services France Grilles, est accessible pour la réalisation des activités scientifiques de l'ensemble des partenaires du GIS, ainsi qu'aux organismes et entreprises ayant des projets communs avec eux.

Pour utiliser le service FG-Cloud, il faut prendre contact avec l'équipe de pilotage.¹⁶ Puis, il faut compléter un questionnaire permettant d'évaluer les besoins et leur adéquation avec le service proposé. La demande est ensuite traitée lors d'une réunion pour l'attribution de ressources. Les sites proposant des ressources au projet fournissent ainsi un accueil adapté à chaque utilisateur.

L'utilisation des ressources FG-Cloud n'est pas facturée. En effet, les infrastructures et leur exploitation sont financées par différentes sources (participation des organismes, subventions régionales, projets, Infrastructure de Recherche France Grilles, etc). Toutefois, si les besoins de l'utilisateur deviennent trop

7. <https://docs.openstack.org/horizon/latest/user/index.html>

8. <https://slipstream.lal.in2p3.fr/>

9. <http://diracgrid.org/>

10. <http://www.france-grilles.fr/catalogue-de-services/fg-irods/>

11. <https://www.egi.eu/>

12. <https://www.egi.eu/services/cloud-compute/>

13. <http://hpc.univ-lille1.fr/>

14. <http://www.grand-est.fr/>

15. <https://www.france-bioinformatique.fr/>

16. <http://www.france-grilles.fr/catalogue-de-services/fg-cloud/>

importants, ils pourraient ne plus être satisfaits par les ressources disponibles qui sont apportées par les sites participants. Dans ce cas, un achat d'équipement (en mutualisation) pourra être envisagé par cet utilisateur (ou équipe).

Afin de justifier la pertinence scientifique de son infrastructure et de ses services, France Grilles demande aux utilisateurs :

- d'ajouter une mention à France Grilles dans la section « remerciements » de leurs publications issues de l'utilisation de ce service ;¹⁷
- de référencer ces publications dans les archives ouvertes HAL¹⁸ et de les ajouter à la collection France Grilles.

Les utilisateurs sont également invités à présenter leurs travaux lors de conférences (journées SUCCES, EGI User Forum, ...).

4.2 Exemples d'utilisation

4.2.1 Cloud challenge

Lors du lancement du service FG-Cloud et pour valider son bon fonctionnement, un appel à propositions a été lancé en 2014. Le projet d'un utilisateur a été sélectionné. Il a pu accéder à l'ensemble des ressources pour effectuer des simulations Monte Carlo massivement parallèles (chimie quantique). Cette première expérimentation a permis de valider l'utilisation de ressources Cloud pour réaliser des calculs scientifiques et nous a permis d'améliorer le service fourni aux utilisateurs.

L'utilisateur a effectué une conférence sur ce challenge lors de l'édition 2015 des journées SUCCES.[1]

4.2.2 Astrophysique

Le service FG-Cloud est régulièrement utilisé par le laboratoire astroparticule et cosmologie¹⁹ (APC). Les ressources *Cloud* sont utilisées pour des travaux de recherche portant sur l'étude de la matière noire et les ondes gravitationnelles. Dans ce cadre, des développements sont en cours pour améliorer le support des conteneurs Docker²⁰ sur l'infrastructure *Cloud*. [2][3][4]

4.2.3 Imagerie médicale

Le laboratoire CREATIS utilise l'infrastructure *Cloud* à travers DIRAC pour exécuter des logiciels d'imagerie biomédicale. Ces outils sont proposés à la communauté biomédicale à travers le portail VIP.²¹ Des évolutions sont actuellement en cours pour utiliser Docker.[5]

5 Conclusion

Le service FG-Cloud est en production depuis maintenant trois ans. De nouveaux sites vont rejoindre prochainement l'infrastructure (GRICAD,²² IRFU,²³ LPNHE²⁴ et le mésocentre de l'Université de Clermont-Ferrand²⁵) et permettre à des utilisateurs toujours plus nombreux de réaliser leurs activités scientifiques sur un *Cloud* de confiance. Afin d'anticiper ces futures évolutions, de nouveaux développements sont en cours, notamment en ce qui concerne l'amélioration de l'authentification centralisée (utilisation de fédérations d'identités existantes), la réplique des machines virtuelles entre les sites (augmentation de la disponibilité des services) ou l'amélioration de l'interface utilisateur (visualisation dynamique des ressources disponibles, ...). Enfin, les efforts sont maintenus pour accompagner les sites à rejoindre la fédération européenne EGI et les utilisateurs dans leur utilisation

17. <http://www.france-grilles.fr/fg-succes/collection-publications/>

18. Hyper Articles en Ligne - <https://hal.archives-ouvertes.fr/>

19. <http://www.apc.univ-paris7.fr/>

20. [https://fr.wikipedia.org/wiki/Docker_\(logiciel\)](https://fr.wikipedia.org/wiki/Docker_(logiciel))

21. <https://vip.creatis.insa-lyon.fr/>

22. Grenoble Alpe Recherche - Infrastructure de Calcul Intensif et de Données - <https://gricad.univ-grenoble-alpes.fr>

23. Institut de Recherche sur les lois Fondamentales de l'Univers - <http://irfu.cea.fr>

24. Laboratoire de Physique Nucléaire et de Hautes Énergies - <https://lpnhe.in2p3.fr/>

25. <https://crri.clermont-universite.fr/Mesocentre-Clermont-Auvergne>

quotidienne des infrastructures.

Bibliographie

- [1] Anthony Scemama. Retour sur le Cloud Challenge France Grilles. Dans actes des Journées SUCCES, Paris, Novembre 2015. https://succes2015.sciencesconf.org/conference/succes2015/AnthonyScemama_SUCCES2015.pdf
- [2] Rémi Fahed et Cécile Cavet. Utilisation du Cloud dans Euclid. Dans actes des Journées SUCCES 2017, Grenoble, Octobre 2017. <https://succes2017.sciencesconf.org/resource/page/id/8>
- [3] Cécile Cavet. Docker for Space Missions. Dans actes de la conférence EGI, Catane, Mai 2017. <https://indico.egi.eu/indico/event/3249/contribution/0/material/poster/0.pdf>
- [4] Cécile Cavet, Antoine Petiteau, Maude Le Jeune, Eric Plagnol, Etienne Marin-Martholaz, Jean-Baptiste Bayle. A proto-Data Processing Center for LISA. Dans les actes du 11ème symposium LISA, Journal of Physics : Conference Series, Volume 840, conférence 1, Zurich, 2016. <http://iopscience.iop.org/article/10.1088/1742-6596/840/1/012045>
- [5] Sorina Pop et Tristan Glatard. DIRAC for Biomed Applications. Dans actes du 6ème DIRAC Workshop, Montpellier, Mai 2016. <https://indico.cern.ch/event/477578>