

Prospective nationale sur les Grilles de calcul

Rapport du groupe de travail **Sciences de la Planète et Sciences de l'Univers**

3 octobre 2008

Monique Petitdidier	IPSL/CETP
Franck Le Petit	LUTH / Observatoire de Paris
Sophie Godin-Beekmann	IPSL/SA
Pierre Le Sidaner	Observatoire de Paris
Jean-Pierre Vilotte	IPGP
Stratis Manoussis	INSU
Geneviève Moguilny	IPGP
Karim Ramage	IPSL
David Weissenbach	IPGP

1- Introduction

Les deux communautés, Sciences de la Planète et Sciences de l'Univers, ont de nombreux points communs. Toutes deux étudient des phénomènes naturels sur lesquels l'Homme ne maîtrise aucun paramètre (météorologie, éruption volcanique, tremblement de terre, évolution des galaxies, ...). Les deux principaux moyens d'investigation sont leur observation et la simulation numérique.

Dans ces communautés, les observations sont précieuses soit par leur caractère unique dû à des événements exceptionnels comme les éruptions volcaniques, explosion de supernovae, et aussi à des conditions atmosphériques, océaniques et climatiques voisines mais toujours différentes, soit par le coût qu'elles représentent. En effet, les missions permettant l'acquisition des données se chiffrent en milliers jusqu'à des centaines de millions d'euros (mission satellites, grands télescopes, mission sur le terrain avec déploiement d'instruments sol et aéroportés). Le volume des données augmente drastiquement en raison de plus grandes résolutions des observations et des simulations, et du déploiement de nombreux réseaux d'observations. Il s'agit de Petaoctets pour les données des modèles climatiques, pour des expériences de missions satellitaires... Par conséquent, les deux communautés ont besoin d'archiver les données acquises, d'y accéder régulièrement et de les exploiter avec des ressources de calcul adaptées au volume de données et aux algorithmes utilisés.

Les simulations numériques sont également au cœur de la recherche. Elles constituent des laboratoires *in silico* pour étudier les processus physiques. Les besoins en moyens de calcul pour les deux disciplines vont de l'ordinateur individuel au super-calculateur en fonction du domaine étudié. Les communautés Sciences de la Planète et Sciences de l'Univers sont parmi les plus importants consommateurs de ressources informatiques sur les grands centres de

calcul nationaux. Les mesocentres associés pour la plupart aux OSU (Observatoires des Sciences de l'Univers) rassemblent des ressources de calcul très utilisées par ces communautés pour les simulations et l'exploitation des données.

Dans les deux communautés, les grands projets se font souvent entre équipes géographiquement dispersées. Ils mettent en relation des équipes de recherche publiques, parfois privées (en particulier en Science de la Planète) ainsi que des agences et des organismes (Météo France, CNES, ESA, NASA ...). La durée de vie des projets est typiquement de quelques années à quelques dizaines d'années dans le cas de missions satellitaires.

Un point clef des deux communautés est leur hétérogénéité et leur interdisciplinarité. Les équipes de recherche en Sciences de la Planète et en Sciences de l'Univers sont spécialisées sur des thématiques particulières : Atmosphères, Océan, Terre d'un côté, Cosmologie, Galaxies, Physique stellaire, Milieu interstellaire, Planétologie de l'autre. Les recherches scientifiques dans les deux communautés font appel à de nombreux domaines de la physique. Les projets rassemblent des équipes de ces communautés avec des expertises complémentaires mais également des équipes d'autres disciplines. Le vaste projet d'observation AMMA par exemple a rassemblé autour de Météo France des équipes travaillant sur l'océan, l'atmosphère, l'hydrologie, l'agriculture, la santé ... De ce morcellement en thématiques et en sujets de recherche, il résulte que les moyens, les outils numériques nécessaires aux équipes de recherche de ces deux disciplines pourront être très différents d'une équipe à l'autre, et d'une phase de leur projet à l'autre.

Malgré cette impression d'éparpillement, ces communautés sont organisées au niveau international depuis des décennies par discipline ou domaine, ont défini des standards en particulier pour les formats et échanges de données, développé des Centres de Données et même dans quelques cas des architectures basées sur des services Web pour accéder aux données et les traiter localement et/ou les télécharger.

Face à cette diversité de besoins des équipes scientifiques, les Grilles de calcul ne peuvent donc constituer qu'un moyen informatique parmi d'autres. De plus, elles doivent être suffisamment souples en terme de *middleware* et d'infrastructure pour accueillir un grand nombre d'applications très différentes et prendre en compte leur environnement et les outils utilisés.

2. – Grilles de calcul

En France, quelques équipes de la communauté Science de la planète disposent déjà d'une forte expérience du calcul sur Grille. L'IPSL s'est impliqué dès 2000 dans DataGrid, premier projet Européen de Grille, suivi par l'IPGP. Puis l'IPSL et l'IPGP ont participé aux différentes phases du projet Européen EGEE et au projet Européen DEGREE (*Dissemination and exploitation of Grids in Earth Science*). Les conclusions de ce dernier projet sur l'adoption de la Grille par cette communauté figurent dans un livre blanc qui indique également les différentes étapes à franchir dans les prochaines années :

http://www.eu-degree.eu/DEGREE/internal-section/wp6/DEGREE-D6.1.2_v2.8.pdf/view

Dans la communauté A&A, les Grilles de production sont encore peu utilisées mais suscitent un intérêt pour certains types de problèmes. Le premier point d'entrée à EGEE en astrophysique vient d'être installé à l'Observatoire de Paris.

3 – Actions pour la prospective de l’Institut des Grilles

La prospective de l’Institut des Grilles fait suite à la prospective sur les super-calculateurs et à celle sur les mesocentres.

Un groupe de travail a été mis en place qui, pour recenser les besoins, les attentes et les verrous en terme de technologie grille dans les communautés Sciences de la Planète et Sciences de l’Univers, a organisé une réunion d’information le 27 mai 2008 à l’Observatoire de Paris et a diffusé un questionnaire à l’attention de l’ensemble des personnels des deux communautés.

3.1 – Réunion d’information

La réunion d’information du mois de mai 2008 a permis de rassembler une trentaine de représentants des deux disciplines ainsi que des membres de la communauté Grille. Il s’agissait de présenter la Grille EGEE, des applications en Sciences de la Planète et Univers, des interfaces pour faciliter l’accès aux utilisateurs ou le portage d’application (Geocluster, Ganga, gEclipse) et les perspectives du ministère et des universités. Une trentaine de représentants de différentes disciplines en Ile de France ont participé à cette réunion.

3.2 – Enquête nationale

Afin de faire un état des lieux sur l’utilisation et le potentiel des Grilles de production dans les deux communautés, un questionnaire a été diffusé à l’ensemble des membres des Sciences de la Planète et Sciences de l’Univers. Des contacts individuels ont également été pris avec des représentants de certaines agences ou thématiques. L’objectif était de toucher directement les responsables et les scientifiques utilisant ou ayant un intérêt dans l’utilisation des Grilles de production pour leur travail de recherche. Des réponses ont été obtenues de tous les principaux établissements de recherche (voir annexe).

Le nombre de réponses est de l’ordre de 200. La majeure partie des laboratoires, instituts et observatoires a répondu. Certaines réponses provenant de responsables d’équipes, de centres de calcul ou de laboratoires, il est difficile de fournir un chiffre pour le nombre de personnes intéressées qui est plus important que le nombre de réponses au questionnaire. Malgré ces réponses globales et une faible connaissance des Grilles de production dans nos disciplines, le nombre de réponses permet de dresser un panorama de l’utilisation et des points de blocage de la Grille dans les deux communautés.

Les questions et les réponses au questionnaire peuvent être trouvées en annexe. Les principaux points que l’on peut extraire sont :

- Une importante fraction des personnes ayant répondu est actuellement limitée par les moyens de calcul et de stockage et par le personnel d’assistance.
- Parmi celles-ci plus de 80% disposent d’applications nécessitant de répéter le même job avec des paramètres différents et 66% ont besoin de partager des masses de données avec des collègues distants.
- 63% des personnes ayant répondu pensent que la grille peut leur être utile mais ne l’ont pas encore utilisée. Et seuls 30% utilisent effectivement la Grille.
- Des personnes ayant répondu, 54% disposent d’applications parallélisées (70% MPI, 30% OpenMP) et 40% requièrent des logiciels sous licences.
- Un tiers des personnes ayant répondu disent être correctement informées sur l’utilité de la Grille pour leurs applications, 8% se disent experts et 59% un peu informées. Près de 80% assisteraient à un séminaire et 73% à une formation de quelques jours à l’utilisation des Grilles.

4 – Utilité de la Grille, besoins & verrous

Les moyens de calcul et de stockage actuels ne sont pas suffisants aux communautés des Sciences de la Planète et des Sciences de l'Univers. Un nombre significatif des personnes ayant répondu se disent intéressées pour utiliser la Grille afin de répondre à ce manque, de traiter plus rapidement leurs problèmes scientifiques, de pouvoir passer à l'étude de cas plus complexes. A partir des applications déjà portées, on constate que la Grille permet d'obtenir des résultats plus difficiles à obtenir autrement et aussi d'ouvrir de nouveaux champs de recherche en raison des ressources informatiques disponibles. Par conséquent, les communautés Sciences de la Planète et Sciences de l'Univers sont intéressées par cette technologie.

De plus, la Grille apparaît comme un moyen de mutualiser les ressources informatiques au niveau mesocentres thématiques et Centres de Données.

Les principaux verrous pour les utilisateurs sont de deux ordres : l'un est l'investissement en temps pour acquérir la technologie afin de déterminer si la Grille est la solution adaptée à un problème spécifique et ensuite l'utiliser ; l'autre, technique, concerne l'installation sur la Grille d'outils et de logiciels couramment utilisés par les applications.

Face à des moyens de type mesocentres ou clusters locaux, la Grille présente des atouts considérables compte tenu du nombre de ressources disponibles. Mais elle ne répondra aux attentes que si :

- les scientifiques peuvent utiliser simplement cette technologie, ce qui nécessite :
 - **des réunions d'informations** sur les Grilles afin de savoir si elle peut être utile pour leurs applications ;
 - **des sessions de formation** pour apprendre à utiliser la Grille ;
 - **une aide au portage d'applications.** Pour permettre un réel transfert de compétences, le portage doit être fait localement dans les équipes. Cela suppose :
 - d'avoir des spécialistes de la Grille au contact des futurs utilisateurs pour l'acquisition du savoir faire et l'aide sur le long terme ;
 - de former à la technologie Grille des ingénieurs numériciens et scientifiques des laboratoires ;
 - Pour les laboratoires ne disposant pas d'ingénieurs en calcul scientifique, des recrutements sont nécessaires.
 - de disposer des outils courants sur la Grille que sont les compilateurs, bibliothèques, logiciels (parfois commerciaux). Le nombre de logiciels commerciaux cités dans les réponses du questionnaire est relativement limité.
- la Grille est suffisamment **stable dans le temps**. Un certain nombre de scientifiques craignent de devoir passer trop de temps à mettre à jour leurs procédures en fonction de l'évolution du *middleware*.
- une **interface sécurisée et facile à implanter** existe pour **l'accès aux données**.

Concernant ces disciplines elles-mêmes puisqu'elles se sont structurées autour de Centres de Données et ont développé des standards et des protocoles d'accès au niveau international hors de la Grille, l'un des verrous est la nécessité de disposer d'interfaces entre la Grille et les Centres de Données et mesocentres des deux disciplines.

5 – Conclusions & recommandations

La Grille intéresse les communautés Sciences de la Planète et Sciences de l'Univers. Elle constitue un moyen de répondre au manque de ressources informatiques de type mesocentre et peut être considérée comme une solution de mutualisation des ressources.

Cependant, les communautés Sciences de la Planète et Sciences de l'Univers sont des communautés à la fois très hétérogènes et très structurées par discipline et domaine.

En ce qui concerne la Grille cela implique que :

- 1) cette technologie ne peut être qu'un moyen de calcul parmi d'autres. Les deux communautés continueront d'avoir besoin de super-calculateurs et de Centres de Données hors de la Grille.
- 2) elle doit supporter un large éventail d'applications ce qui suppose que les principaux logiciels utilisés dans ces communautés soient disponibles sur la Grille (compilateurs, bibliothèques, principaux logiciels commerciaux).
- 3) Les deux communautés s'étant structurées autour de Centre de Données hors de la Grille, elles n'accepteront la Grille que si des interfaces entre les deux sont développées et que ces interfaces remplissent les conditions d'interopérabilité avec celles définies dans d'autres pays.
- 4) Cette technologie facilite la collaboration entre les équipes au niveau européen et international.

Un verrou important est le temps à investir pour apprendre à utiliser quotidiennement cet outil. Par conséquent, plusieurs recommandations peuvent être faites. Il semble nécessaire d'organiser :

- des réunions d'informations pour aider les scientifiques à identifier les applications qui bénéficieraient d'un portage sur la Grille.
- des tutoriaux de formation à l'attention des scientifiques et des ingénieurs en calcul scientifique des laboratoires.
- des portages concrets d'applications sur la Grille destinées à un groupe d'utilisateurs.
- un support pour interfacer les Centres de données des deux communautés et la Grille.

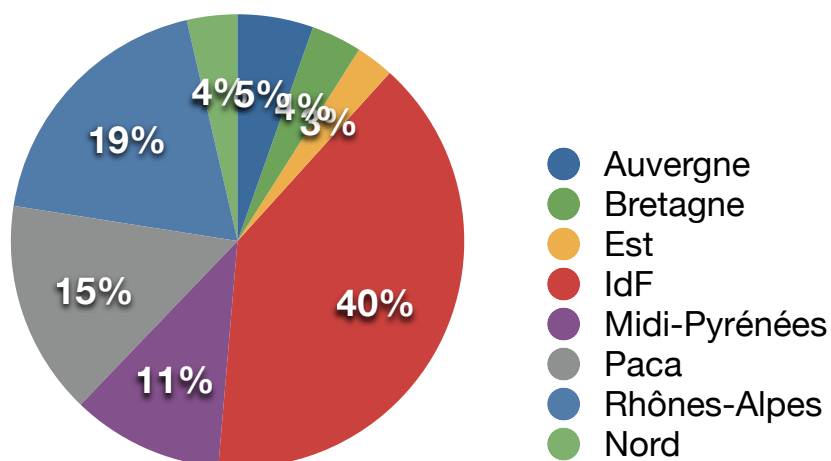
La sous-utilisation actuelle de la Grille en Sciences de la Planète et Sciences de l'Univers provient d'un manque de contact entre les scientifiques de ces communautés et les experts de la Grille. Pour résoudre ce problème, des réunions et tutoriels doivent être organisés pour que la technologie soit acquise au sein des laboratoires. Afin que les bénéfices de ces tutoriels s'inscrivent dans la durée, il est nécessaire de mettre en place un **réseau territorial de spécialistes de la Grille dans les deux communautés**. L'expertise grille se trouvant naturellement là où sont localisés les points d'accès à la Grille, une façon de procéder pourra être de favoriser le déploiement des points d'accès à la Grille à travers le territoire national dans les deux communautés.

Ces conclusions et recommandations faites au niveau national sont très proches de celles des mêmes communautés au niveau Européen. Quelques projets européens et des organisations internationales (Open Geospatial Consortium, Global Earth Observation System of Systems, EuroVO-DCA, Open Grid Forum) cherchent des solutions techniques et standards pour répondre aux besoins de nos communautés c'est-à-dire définir et bâtir des interfaces entre cette technologie de Grille et celles déjà développées par nos communautés. Même si des solutions commencent à exister, elles réclament des moyens humains et financiers pour les mettre en œuvre.

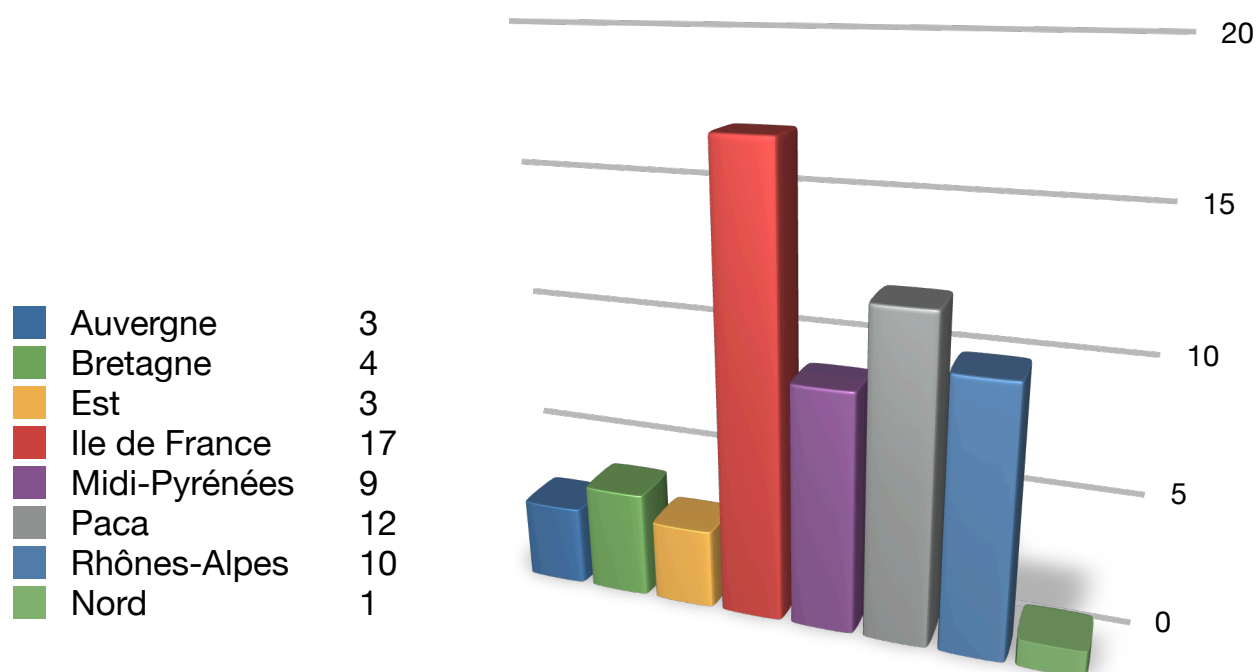
ANNEXE

Résultats de l'enquête

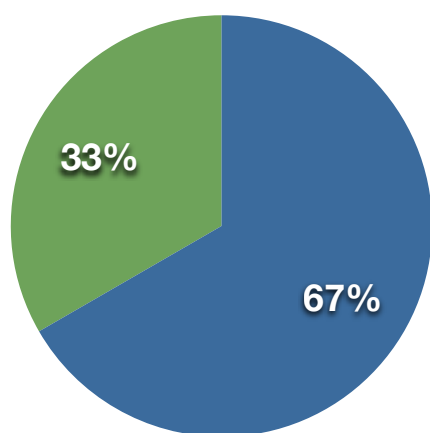
A0 - Localisation géographique des réponses



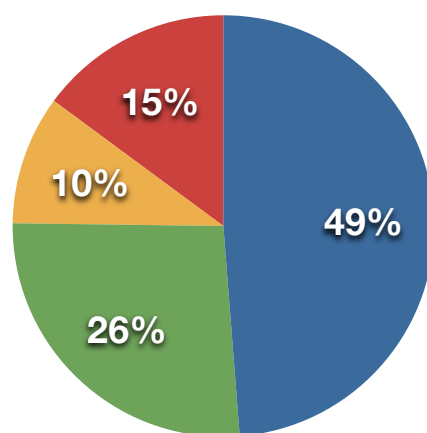
A0 - Nombre de laboratoires ayant répondu par régions



A1 - Identité des personnes ayant répondu



- Individu
- Responsable

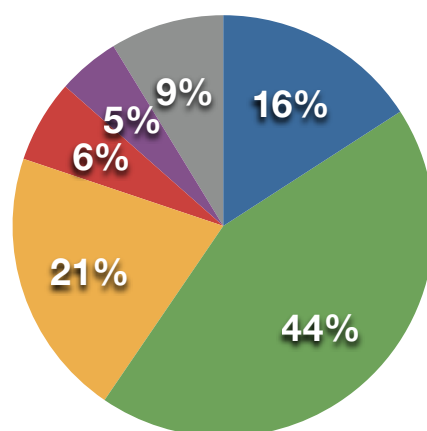


- Chercheur
- Ingénieur
- Doctorant
- Autres

A2 - Limitations en ressources dans le contexte actuel

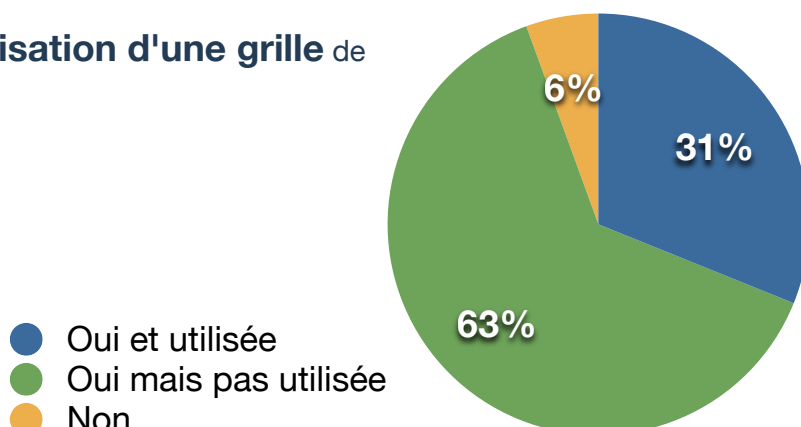
Etes-vous **limité par les ressources informatiques** actuellement à votre disposition pour mener à bien vos recherches ?

- non
- Oui - Calcul
- Oui - Stockage
- Oui- Réseau
- Oui - Services
- Autres



A3 - Intérêt de la Grille

Verriez-vous un **intérêt à l'utilisation d'une grille** de production ?



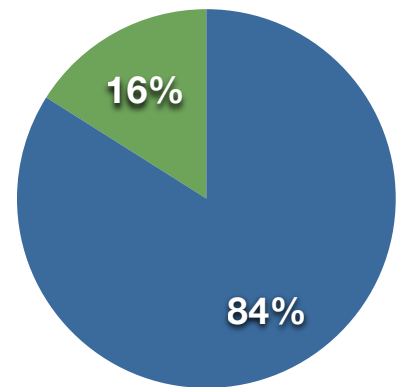
Pensez-vous que cette technologie et les moyens informatiques associés auraient une **utilité** pour vous ?



A4 - Types d'applications

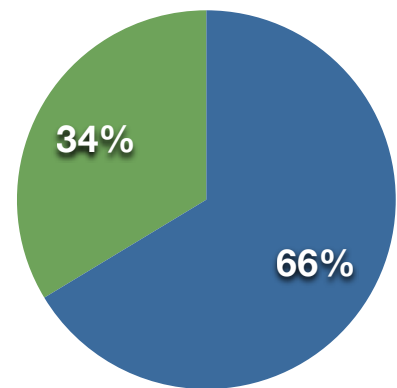
Travaillez-vous avec des applications dans lesquelles vous devez **exécuter un grand nombre de fois un programme** en faisant varier ses paramètres ?

● Oui
● Non



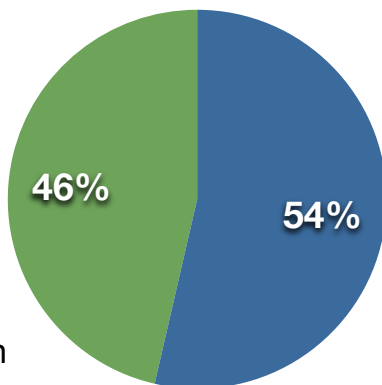
Avez-vous à utiliser ou produire de grandes masses de données ou appartenez-vous à des projets impliquant un partage de **grandes masses de données** avec des collègues distants ?

● Oui
● Non

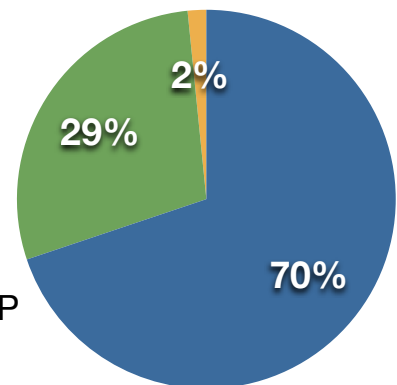


Vos applications sont-elles **parallélisées** ?

● Oui
● Non



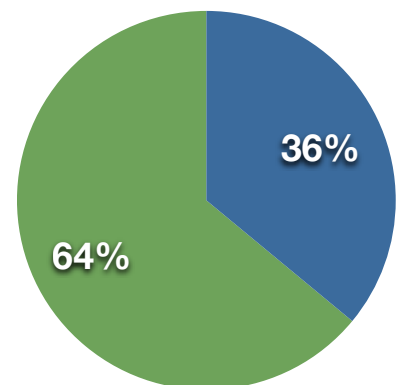
● MPI
● OpenMP
● Autre



Vos applications utilisent-elles un **logiciel sous licence** ?

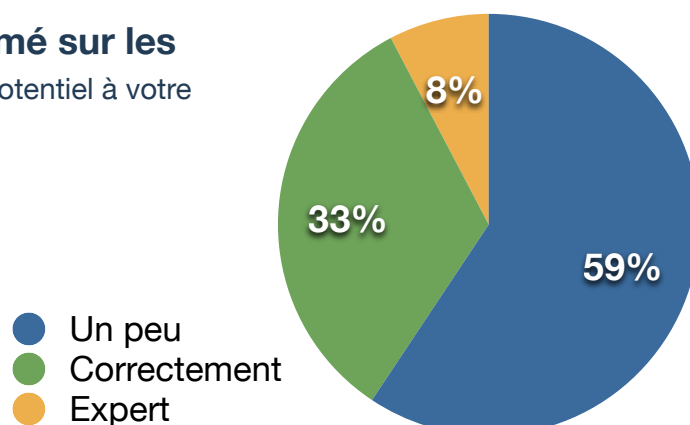
IDL	Codior	NAG
Matlab	ifort	VASP
Geocluster	FLAC 3D	VU
Molpro	LGPL	Envi

● Oui
● Non

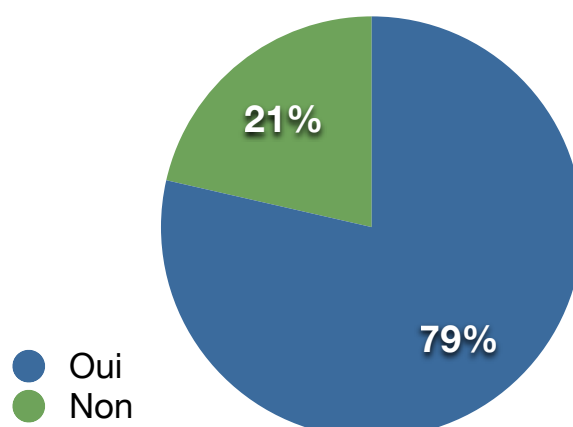


A5 - Connaissance de la Grille & Formation

Pensez-vous être **suffisamment informé sur les grilles de production** et leur apport potentiel à votre recherche ?



Souhaiteriez-vous assister à un **séminaire** d'introduction aux grilles de production ?



Etes-vous prêt à suivre une **formation de quelques jours** sur l'utilisation des grilles de production ?

