

High Throughput Computing (HTC)

Pour beaucoup de scientifiques, le progrès et la qualité scientifiques de la recherche sont fortement liés au débit de calcul. En d'autres termes, la plupart des scientifiques sont concernés par combien d'opération à virgule flottante par mois ou par année ils peuvent extraire à partir de leur environnement de calcul plutôt que le nombre de telles opérations l'environnement peut fournir par seconde ou minute.

Les opérations à virgule flottante par seconde (FLOPS) ont été la mesure employée pour l'évaluation des environnements de Calcul Haute Performance (HPC).

Peu d'attention a été consacrée par la communauté informatique aux environnements qui peuvent fournir de grandes quantités de capacité de traitement sur de longues périodes. Nous faisons référence aux environnements de Calcul Haut Débit (HTC : High Throughput Computing).

Le traitement à haut débit (HTC) est opposé au traitement à haute performance (HPC). Là où le premier est évalué sur le long terme, le second implique des résultats à court terme. C'est ainsi que plutôt qu'en MIPS (millions of instructions per second), les performances des applications à haut débit sont mesurées en TIPY (trillions of instructions per year).

Calcul Haut débit (HTC)	Calcul Haute Performance (HPC)
<ul style="list-style-type: none">• Traitement des calculs à « long terme »• La puissance de milliers de processeurs• Le débit est le principal but• L'ordre dans lequel les différentes tâches s'exécutent n'est pas important• Si l'équipement est indisponible, le travail est migré sur une autre machine	<ul style="list-style-type: none">• Traitement des calculs à « court terme »• Utilisation d'un unique processeur coûteux• Si un processus échoue, le calcul s'arrête• Tous les processus doivent être redémarrés au même point de synchronisation

La clef à HTC est la gestion et l'exploitation efficaces de toutes les ressources informatiques disponibles. Puisque les besoins de calcul de la plupart des scientifiques peuvent être satisfaits de nos jours par le CPU et la mémoire de l'ordinateur, le rendement élevé ne joue pas un rôle important dans un environnement de HTC. Le plus grand challenge d'un environnement HTC est comment maximiser la quantité de ressources accessibles aux utilisateurs.

En d'autres termes, le but ici n'est pas de rechercher la performance pure comme dans le cas précédent mais de rentabiliser au maximum des ressources en récupérant les cycles processeurs non utilisés quand celui-ci est oisif (*idle*). On pense bien entendu aux stations de travail qui sont généralement très largement sous utilisées, ne serait-ce qu'en dehors des heures de bureau.

II. Les grilles informatiques

Depuis quelques années, l'informatique répartie et les technologies associées sont dans une constante évolution technologique et économique. Dans ce contexte, le terme « The Grid » ou « grille de calcul », a été introduit pour décrire une infrastructure de calcul répartie utilisée dans des projets de recherche scientifique et industrielle aux Etats-Unis durant les années 1990.

a. Pourquoi les grilles informatiques ?

De plus en plus les applications développées de nos jours sont gourmandes en puissance et en stockage. Ce qui entraîne un problème de dimensionnement des ressources.

Aussi ces applications utilisées par une communauté géographiquement distribuée permet le travail collaboratif. D'où le partage des données et des applicatifs posant ainsi le problème d'accès partagé aux ressources.

Pour les chercheurs, il fallait trouver des possibilités pour accéder aux ressources, interagir avec des collègues, analyser d'énormes volumes de données et de partager des résultats. Ensuite incorporer des ressources existantes qui sont hétérogènes et dynamiques.

L'accès aux services informatiques devrait être comparable en coût et facilité au branchement d'appareils sur le réseau électrique. Cet accès devrait être aisé et transparent quels que soient la puissance nécessaire, la complexité des équipements matériels et logiciels mis en oeuvre, les fournisseurs du service.

Raisons des grilles de calcul

Il est important de savoir quels avantages une grille de calcul est en mesure d'offrir que les infrastructures et les technologies actuelles ne sont pas capables d'assurer. Nous exposons dans la suite quelques unes des raisons pouvant amener à déployer une grille de calcul.

- **Exploiter les ressources sous utilisées** : les études montrent que les ordinateurs personnels et les stations de travail sont inactifs la plupart du temps. Le taux d'utilisation varie entre 30% (milieux académiques et industriels) et 5% (machines grand public). Les grilles de calcul permettront ainsi d'utiliser les cycles processeurs durant lesquels les machines sont inactives afin de faire tourner une application nécessitant une puissance de calcul importante et que les machines qui lui sont dédiées n'arrivent pas à assurer. Les cycles processeur ne sont pas la seule ressource sous utilisée ; souvent les capacités de stockage le sont aussi. Ainsi il est possible qu'une grille agrège toutes ces ressources afin de les partager entre les différents utilisateurs

(on parle alors de « grille de données » ou Data Grid). Une autre conséquence d'une telle utilisation est la possibilité de faire du partage de charge entre les différentes ressources d'une grille.

- **Fournir une importante capacité de calcul parallèle** : le fait de pouvoir fournir une importante capacité de calcul parallèle constitue une caractéristique importante des grilles de calcul. En plus du domaine académique, le milieu industriel bénéficiera énormément d'une telle capacité : bioinformatique, exploration pétrolière, industrie cinématographique etc. en effet les applications sont écrites d'une façon à pouvoir exploiter parallèlement des ressources (clusters, machines multiprocesseur ...). Les grilles de calcul peuvent de la même manière fournir des ressources dont l'utilisation pourra se faire en parallèle.
- **Meilleure utilisation de certaines ressources** : en partageant les ressources, une grille pourra fournir l'accès à des ressources spéciales comme des équipements spécifiques (microscope électronique, bras robotique ...) ou des logiciels dont le prix de la licence est élevé. Ainsi ces ressources exposées à tous les utilisateurs seront mieux utilisées et partagées et ainsi on évitera d'avoir recours à installer du nouveau matériel ou acheter de nouvelles licences.
- **Fiabilité et disponibilité des services** : du fait que les ressources fédérées par une grille de calcul soient géographiquement dispersées et disponibles en importantes quantités permet d'assurer la continuité du service si certaines ressources deviennent inaccessibles. Les logiciels de contrôle et de gestion de la grille seront en mesure de soumettre la demande de calcul à d'autres ressources.

b. Définition

La grille informatique (ou Grid Computing) a été nommée ainsi par analogie au réseau électrique (ou Electric Power Grid) qui permet d'obtenir une puissance électrique grâce à une interface simple (la prise de courant) sans se soucier de la provenance de cette énergie et de la manière dont elle a été fabriquée.

La grille reprend ce principe : plusieurs ordinateurs alimentent un réseau pour fournir une puissance globale. Chaque utilisateur dispose alors de ressources considérables, aussi bien en terme de puissance de calcul que de stockage de données, au moyen d'une interface simplifiée.

On peut définir aussi les grilles informatiques comme un système réparti haute performance dont certains des constituants sont des calculateurs parallèles ou des grappes.

Ce concept était à la base essentiellement technique, avec pour objectif principal la mise à disposition de ressources de traitement informatique très importantes et multi organismes avec un accès aussi transparent que possible. Le concept a fait ses preuves autour d'infrastructures prototypes; des grilles de production sont maintenant opérationnelles en entreprises et entre organismes. Le déploiement d'infrastructures de ce type en production a des impacts forts sur le plan des réseaux, de la sécurité, de l'administration au sein de chaque organisme, et sur le plan organisationnel puisqu'un l'un des objectifs visés est celui du partage de ressources entre entreprises ou institutions.

En effet, au delà des besoins en capacité de traitement, l'approche grille permet la co-allocation et le partage dynamique de toutes les ressources de systèmes hétérogènes indépendamment des unités organisationnelles (départements d'entreprises, institutions et entreprises). Des organisations virtuelles (VO) permettent maintenant de définir en tant que de besoin des communautés d'utilisateurs partageant suivant des modalités communes d'utilisation, des sous-ensembles de ces ressources globalisées de traitement, de stockage, des logiciels, des instruments et des services largement distribuées, quelles que soient les organisations propriétaires.

Un nouveau moteur de la puissance informatique

Depuis quelques mois, la grille informatique, fait de plus en plus parler de lui dans l'univers des nouvelles technologies. Il se fait aussi progressivement connaître du grand public à travers des expériences scientifiques et industrielles largement relayées par les médias. Cependant, il reste souvent difficile de comprendre ce qui se cache derrière cette technologie, dont les premières expériences ont débuté il y a une trentaine d'années.

De manière plus formelle, la technologie de grille informatique est une forme d'informatique distribuée, basée sur le partage dynamique des ressources entre des participants, des organisations et des entreprises dans le but de pouvoir les mutualiser, et faire ainsi exécuter des applications de calcul intensif ou des traitements de très gros volumes de données. Ces applications ne seraient pas possibles au sein d'un seul organisme ou d'une seule entreprise. Du point de vue de l'utilisateur, l'origine des ressources utilisées est totalement abstraite et transparente. Ce dernier ne se soucie de rien : ni de la puissance nécessaire pour ses

applications, ni du type de machines utilisées, et encore moins de leur localisation physique. C'est la plate-forme de Grid qui gère tous ces aspects.

En pratique, le Grid Computing essaye d'aller au-delà de l'Internet. Les acteurs impliqués dans ce projet ont un rêve : celui de l'utilisation à la demande. Son objectif est de rendre disponibles toutes les ressources de calcul et de stockage des ordinateurs connectés sur une grille informatique pour des applications de gestion, de communication, de business ou encore des applications typiquement scientifiques.

Il n'y a aucune raison pour que ce qui est possible pour des entreprises ou des centres de recherche ne le soit pas également pour nos collectivités locales. Celles-ci possèdent en effet des parcs informatiques importants et demain elles auront besoin d'une puissance informatique décuplée pour pouvoir utiliser les nouveaux outils de réalité virtuelle et de simulation en 3D. Nos collectivités ont donc tout à fait intérêt à organiser leurs parcs informatiques en grille informatique pour pouvoir bénéficier, sans augmentation de coût, d'une puissance de calcul correspondant à leurs futurs besoins. [2]

Une tentative de classification [3]

○ **Classification par objectifs**

Grille d'information: partager la connaissance

Le web

Observatoires virtuels (<http://www.france-ov.org/>)

Grille de stockage: stocker à grande échelle

Observatoires virtuels

LCG (15 Petabytes/an) (<http://lcg.web.cern.ch/LCG/>)

Grille de calcul: agréger la puissance de calcul

EGEE (<http://www.eu-egee.org/>)

DEISA (<https://www.deisa.org/>)

CiGri (<https://ciment.imag.fr/cigri>)

SETI@home (<http://setiathome.berkeley.edu/>)

Grille d'expérimentation: Recherche en informatique distribuée

Grid5000 (<http://www.grid5000.fr>)

PlanetLab (<http://www.planet-lab.org/>)

DAS3 (<http://www.cs.vu.nl/das3/>)

XtremLab (<http://xtremlab.lri.fr/>)

Grille « best-effort »: Exploiter les temps morts des clusters au sein d'une communauté dont les ressources sont utilisées de manière discontinue dans le temps

○ **Classification par infrastructure**

Grilles d'institutions

(EGEE, Grid5000, CiGri,...)

- Grilles de grappes situées dans des centres de calcul
- Ordre de grandeur: centaines de noeuds par grappe
- Noeuds stables et relativement sécurisés

Desktop computing, systèmes Peer-to-peer

(Seti@Home, Boinc, XtremLab,...)

- PC sur Internet
- Ordre de grandeur: millions de noeuds disséminés
- Noeuds très volatiles et peu sécurisés

c. Applications des grilles de calcul

Les principales applications des grilles de calcul se feront dans les domaines suivants :

- **Supercalculateur réparti** (*Distributed Supercomputing*) : une grille de calcul pourra agréger une importante quantité de ressources afin de fournir la puissance de calcul nécessaire pour de nombreuses applications et que même les supercalculateur les plus modernes ne peuvent fournir. Selon la nature et l'étendue de la grille, les ressources agrégées pourront aller de toutes les stations de travail d'une université à tous les supercalculateurs des organismes de recherche d'un pays. Comme exemples d'applications nous pouvons citer la simulation distribuée dans la météorologie, la cosmologie et l'aéronautique ...
- **Calcul haut-débit** (*High-Throughput Computing*) : une grille de calcul sera utilisée pour ordonnancer en parallèle une importante quantité de tâches indépendantes les unes des autres. Comme exemples d'applications nous pouvons citer la recherche de clés cryptographiques, les simulations de molécules et l'analyse du génome ...
- **Calcul sur demande** (*On-Demand Computing*) : une grille de calcul pourra fournir les ressources pour satisfaire les demandes à court terme d'une application et que les ressources locales ne sont pas en mesure d'assurer (cycles processeur, stockage ...). Le défi principal pour les concepteurs de telles grilles est la nature dynamique et aléatoire des demandes faites par les utilisateurs qui peuvent constituer une large population.
- **Calcul Collaboratif** (*Collaborative Computing*) : cette classe d'applications inclut les applications d'interaction entre humains dans des environnements de simulation en temps-réel (conception et interaction avec un nouveau moteur d'avion, conception d'un plan urbain ...).

- **Génération, traitement et stockage d'énormes quantités de données** (*Data-intensive Computing*) : dans de telles applications, une grille de calcul pourra absorber et stocker l'importante quantité d'information générées. Comme exemple d'applications nous pouvons mentionner la production d'une carte de l'univers, la prévision météorologique à long terme, les simulations quantiques ...

d. Architecture générale d'une grille informatique

Bien que chaque projet ait sa propre architecture, une architecture générale est importante pour expliquer certains concepts fondamentaux des grilles. C'est ainsi que l'architecture d'une grille peut être représentée sous forme de couches, chacune assurant une fonction spécifique.

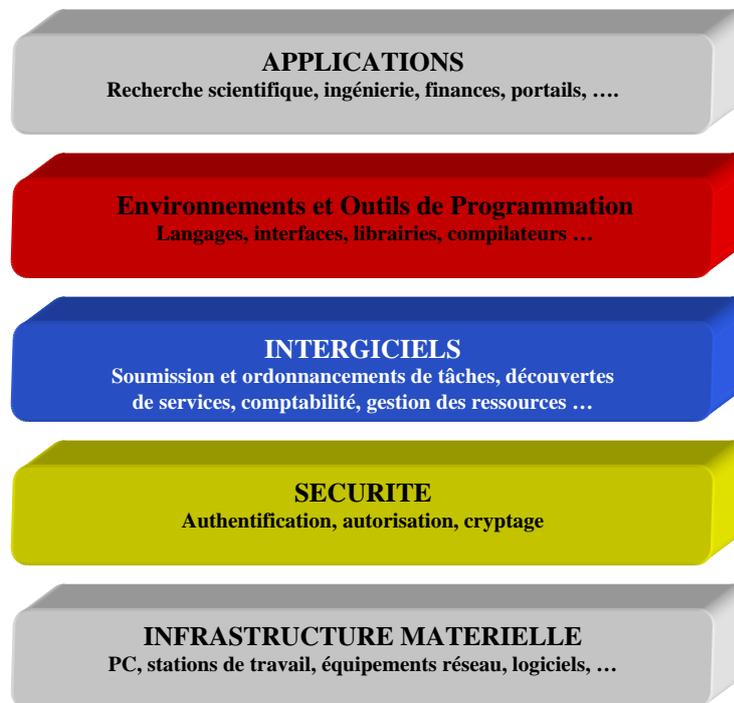


Figure 1 : Architecture en couches des grilles informatiques

Au fond de l'architecture en couches nous avons les ressources proprement dites interconnectées à travers des réseaux locaux et grande distance. Cette **infrastructure matérielle** (fabric) comprend notamment des PC, des stations de travail, des grappes de calcul (clusters), des équipements de stockage, des bases de données, des équipements spéciaux ...

Les ressources d'une grille peuvent varier des plus simples (machines de même architecture, même système d'exploitation, réseau d'interconnexion homogène) au plus compliquées

(différentes architectures de machines, plusieurs systèmes d'exploitation, étendue géographique, différentes politiques de sécurité, plusieurs administrateurs ...).

La seconde couche fournit les mécanismes de **sécurité** nécessaires à la protection des ressources. Dans les grilles de calcul, la sécurité est un souci beaucoup plus important que dans les environnements d'informatique répartis traditionnels. Le nombre important de ressources et leur étendue géographique constituent un facteur important. Ainsi il est fondamental de bien identifier les ressources à protéger et les degrés de protection nécessaires à appliquer. Parmi les mesures de protection on trouve ceux normalement employés dans les réseaux (pare-feu, détection d'intrusion ...) ainsi que ceux spécifiques aux environnements répartis et aux grilles (authentification, autorisation, single sign-on ...).

La troisième couche fournit les **intergiciels** (*middleware*) nécessaires à la gestion des ressources, la coordination de l'accès, l'ordonnancement des tâches ...

Nous expliquons ci-dessous brièvement les principales fonctions assurées par ces intergiciels:

- **Ordonnancement** (*scheduling*) : un ordonnanceur (scheduler) devra trouver la machine la plus appropriée pour faire tourner les tâches qui lui sont soumises. Les ordonnanceurs peuvent aller du plus simple (allocation de type round-robin) au plus compliqué (ordonnancement à base de priorités ...). Les ordonnanceurs réagissent à la charge de la grille. Ils déterminent le taux de charge de chaque ressource afin de bien ordonner les prochaines tâches. Ils peuvent être organisés en hiérarchie avec certains interagissant directement avec les ressources, et d'autres métaordonnanceurs interagissant avec les ordonnanceurs intermédiaires. Ils peuvent aussi superviser le déroulement d'une tâche jusqu'à sa terminaison, la soumettre à nouveau si elle est brusquement interrompue et la terminer prématurément si elle se trouve dans une boucle infinie d'exécution.
- **Réservation** : il est possible dans les grilles de calcul de réserver les ressources à l'avance et ceci afin de garantir une certaine qualité de service et de respecter certaines échéances. Pour cela l'intergiciel devra fournir un système de réservation permettant aux utilisateurs d'exprimer leurs besoins en ressources et d'effectuer la réservation.
- **Services d'information et d'annuaire** (Information and Directory services) : une grille est un environnement où le nombre et la charge des ressources sont dynamiques. Un objectif lors de la conception d'une grille est de rendre les ressources facilement

accessibles à tous les processus. Il est alors nécessaire de fournir des mécanismes permettant d'enregistrer et de découvrir ces ressources et d'identifier leur état.

- **Service de nom** (Naming Service) : Comme dans tout système réparti, une grille devra permettre de référencer ses ressources d'une façon standard et uniforme.

Une grille de calcul devra fonctionner avec un large spectre de technologies matérielles ou logicielles. Tout comme les utilisateurs du WEB ne se soucient pas de savoir si les serveurs WEB tournent sur une machine x86 ou Mac ou bien si le système d'exploitation est Unix ou Windows, les utilisateurs d'une grille ne veulent pas se préoccuper des détails matériels et logiciels de l'infrastructure. Un environnement de grille idéal fournira ainsi un accès transparent aux ressources en cachant toutes les différences « physiques ».

La quatrième couche regroupe tous les outils et les paradigmes pouvant aider les développeurs à concevoir et écrire des applications pouvant tourner sur la grille. On y trouve plus particulièrement des compilateurs, des bibliothèques, des outils de conception d'applications parallèles ainsi que des interfaces de programmation ou API (découverte et réservation de ressources, mécanismes de sécurité, stockage ...) que les développeurs d'applications pourront utiliser.

La dernière couche regroupe les **applications** proprement dites qui sont de nature variée : projets scientifiques, médicaux, financiers, ingénierie ...

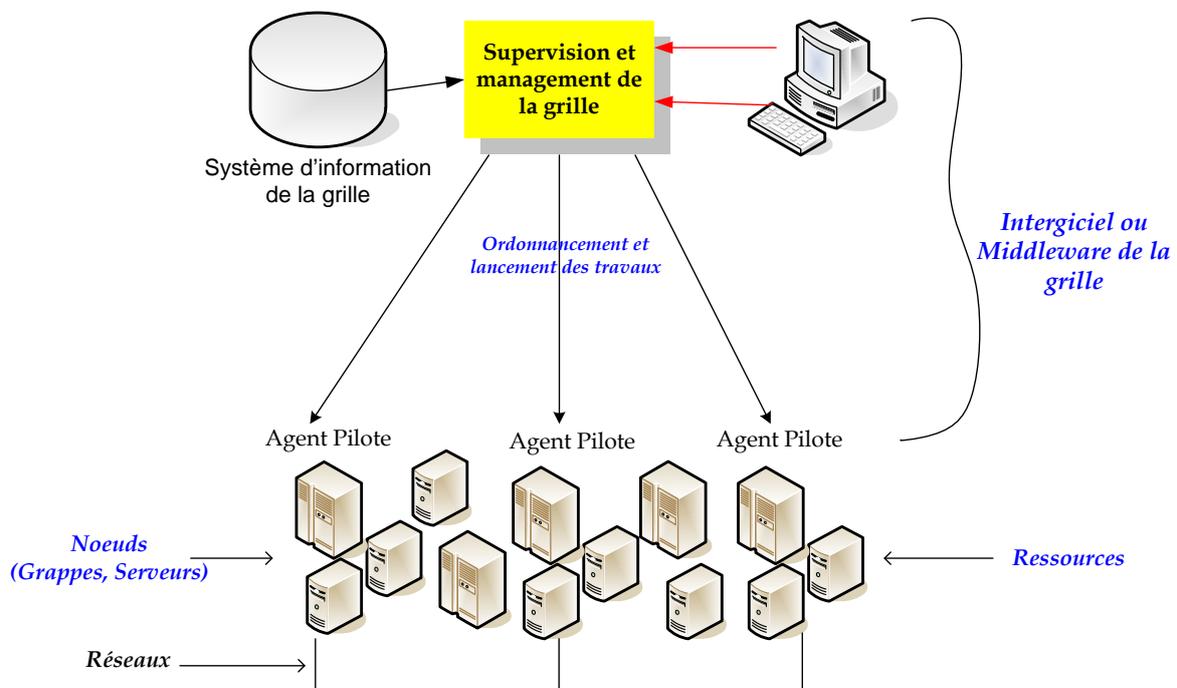


Figure 2 : Architecture simplifiée d'une grille

e. Le paysage des grilles

L'on ne compte plus, aux Etats-Unis tout comme au Japon où l'on trouve par exemple un *Grid technology Research Center* les projets de recherche et développement liés aux grilles, simplement nous noterons que l'effort est mis sur les aspects ingénierie et organisation des architectures matérielles pour aller vers des infrastructures « de production ». Les grilles entrent dans une phase opérationnelle et l'on voit apparaître des organismes fournissant des *Grid and Networking Services*. Le mot d'ordre européen est *Building Grid for Europe a crucial technology for science and industry*, les usages des grilles sont une des priorités affichées du programme *Information Society* de la commission européenne.

Là, le terrain a été défriché par les projets EuroGrid et DataGrid qui ont validé le modèle de grille et qui ont débouché sur les grands projets d'infrastructure Deisa, EGEE et applicatifs BEinGRID.

L'excellence internationalement reconnue de la recherche française (INRIA, Universités, CNRS) a conduit l'INRIA à piloter le réseau d'excellence (NOE) CoreGRID lancé en 2004. Enfin, l'on trouve en Europe des projets tels que e-Science, D-Grid, Grid.it, Grid-Ireland, aux dimensions certes très variables mais qui à chaque fois fédèrent les besoins et les initiatives concernant les grilles au niveau national et dont les acteurs jouent tout naturellement un rôle majeur au niveau européen.

En France, les actions de recherche ont été fédérées au travers des ACI Grid, Grid5000, Le projet e-Toile soutenu par le Réseau National des Technologies Logicielles et piloté par le Centre Nationale de la Recherche Scientifique (CNRS) a permis de valider, une approche grille haute performance pour les applications de la simulation numérique.

Entreprise / Organisme	Produit	Origine	Technologie
Avaki	Legion (Compute & Data)	USA	Mixte
CNRS/LRI	XtremWeb	F	Desktop Grid
CNRS/LAAS+ Alinka SA	Aroma	F	Outils d'administration
CNRS/INRIA/CEA/EDF/CS/SUN	e-Toile	F	Massive Grid
Data Synapse	LiveCluster	USA	Desktop Grid
Entropia	DC Grids	USA	Desktop Grid
GridSystems	Inner Grid	SP	Desktop Grid
GridXpert	Synergie	F -> US	Outils d'administration
Icatis	ComputeMode	F	Extension Cluster>PC
HP	Utility Data Center	USA	Massive Grid (1e gén.)
IBM	Pas de produit propre	USA	
Meshtechnologies	OfficeGrid	DK	Desktop Grid
Pallas->Intel	Unicore	D->USA	Massive Grid (1e gén.)
Platform	LSF/Symphony	USA	Mixte
Sun	Sun Grid Engine	USA	Massive Grid (1e gén.)
Université d'Argonne	Globus	USA	Massive Grid
Université de Berkeley	Boinc	USA	Desktop Grid
Université du Wisconsin	Condor	USA	Massive Grid
United Devices	Grid MP	USA	Desktop Grid
W2G technologies	Web2grid	F	Virtualisation (V V M)

Figure 3 : Quelques fournisseurs et produits à considérer

f. Impacts sur les réseaux

Un gros effort a été accompli sur les performances d'interconnexion locale de grappes de serveurs avec des technologies comme Myrinet¹ ou Infiniband². Les grandes grilles doivent traverser des WAN, avec les problèmes de bande passante et surtout de latence liés, ce qui induit des difficultés dans la prévision des délais d'exécution de travaux et surtout dans l'architecture des applications.

La mise en place d'une infrastructure de calcul distribué telle que la grille a des conséquences sur les réseaux. La plus évidente (et celle attendue par un administrateur) peut être une augmentation de l'utilisation des liens, même si cela est loin d'être automatique et forcément significatif. En effet, les applications « grille » n'engendrent pas toutes des transferts de gros volume de données. Le cas échéant, des infrastructures privées sont envisagées afin d'une part de garantir le service pour les expériences et de ne pas perturber l'utilisation normale des réseaux. L'Internet public ne sera alors utilisé que pour l'accès à la grille, la consultation des données ou l'exécution d'applications spécifiques.

Parmi les conséquences inattendues, citons d'abord les conséquences directes sur tous les types de réseau : réseau local, métropolitain puis régional selon la hiérarchie adoptée dans chaque pays, certains pouvant être absents. La première conséquence réside dans le besoin de gestion opérationnelle. Ce n'est pas tant la haute disponibilité des réseaux de bout-en-bout qui est critique, l'architecture distribuée de la grille s'accommodant relativement bien de l'indisponibilité d'un site. C'est plutôt la remontée et la résolution des incidents qui sont nécessaires. La disponibilité des informations dans un système centralisé à destination des utilisateurs de la grille est un besoin qui a un impact fort sur les réseaux s'ils n'ont pas déjà un système opérationnel en place.

g. Utilisation en entreprise

Conçues à l'origine pour permettre l'exécution d'applications nécessitant de grandes puissances de calcul et utilisées principalement par les chercheurs, les grilles informatiques sortent aujourd'hui des laboratoires et commencent à intéresser fortement le monde l'industrie.

¹ Myrinet est une technologie récente d'interconnexion de réseau, commuté à faible latence et à haute performance, basée sur l'envoi de messages. Les produits d'interconnexion Myrinet permettent de connecter à très haute vitesse un très grand nombre d'ordinateurs afin de former des clusters très rapides.

² Infiniband est un bus d'ordinateur bi-directionnel à faible coût avec un débit de 10 Gbits/s. Il utilise une technologie permettant à plusieurs périphériques d'accéder en même temps au réseau.

En se raccordant à une grille de calcul, les informaticiens transformeront instantanément leur infrastructure de traitement de l'information en un guichet centralisé donnant accès à un nouveau monde – littéralement : le monde entier – de ressources de calcul. Les entreprises pourront immédiatement résoudre les problèmes de leurs clients, donner à leurs collaborateurs répartis partout dans le monde les moyens nécessaires pour leurs transactions commerciales, mais aussi proposer des produits et des services à l'échelle internationale.

Mais le plus intéressant dans ce modèle de traitement, c'est son impact financier. Grâce à la technologie grille de calcul, le budget et les ressources consacrés à l'informatique ne sont plus gaspillés en coûts d'équipement et de gestion. Tout comme les opérateurs des réseaux de distribution d'électricité, d'eau ou de communications téléphoniques, la grille de calcul fournit sa propre infrastructure informatique. Les entreprises paient donc uniquement pour l'utilisation des cycles de calcul, et non pour l'équipement informatique proprement dit.

En entreprise, les grilles se mettent surtout en place dans de très grandes sociétés à haute technologie pour la modélisation, la simulation et la visualisation 3D dans l'automobile, l'aéronautique, le spatial, le recherche pétrolière et le nucléaire. Les autres projets couvrent le calcul intensif et une meilleure gestion du parc informatique avec son obsolescence rapide. Les grilles se généralisent aussi rapidement dans la pharmacie, la biologie, la modélisation moléculaire et la chimie. En finance et bourse, on a pu observer un abaissement drastique des traitements dans l'analyse de risque dans la finance et la bourse. Les grilles ne sont pas réservées aux grandes Entreprises uniquement. Plusieurs fédérations de PME s'y intéressent pour le partage dynamique de leurs équipements, données techniques et codes de calcul, ou en vue d'utilisation en mode informatique à la demande avec des fournisseurs externes, pour des coûts actuellement hors de leur portée. Donc pour les PME, c'est un outil économique de recherche et de développement.

IBM propose une grille informatique pour les PME. Sa nouvelle offre, baptisée « Grid and Grow Express » vise les entreprises de 100 à 1000 utilisateurs. Elle est basée sur sept serveurs lames eServer BladeCenter dont l'entreprise peut adapter la configuration en fonction de ses besoins.

SUN a lancé aussi la grille de calcul « Sun Grid Compute Utility » constitué de 5000 processeurs Opteron et UltraSPARC. Cette grille permet à n'importe quelle entreprise de louer une partie de la grille quelques heures sur Internet afin d'y effectuer un calcul complexe. Elle évite ainsi à l'entreprise d'avoir à investir dans un supercalculateur.

III. LES AGENTS MOBILES

Il existe différents domaines de l'informatique à savoir les systèmes multiagents³, les agents d'interface⁴, les agents mobiles, utilisant le concept d'agent. L'objectif de cette partie est de présenter les travaux touchant à la notion d'agents mobiles en essayant d'en dégager les concepts fondamentaux.

a. Pourquoi la mobilité ?

L'exploitation de la mobilité dans les réseaux ouvre des possibilités très grandes pour améliorer leur fonctionnement général (sécurité, fiabilité, QoS, ...). [5]

Elle permet :

- La réduction de la charge des réseaux : Les systèmes distribués génèrent de nombreux messages pour exécuter leurs tâches. En envoyant les agents où les tâches se font, les messages échangés deviennent locaux et libèrent d'autant la charge des réseaux ; ce qui réduit les délais de communication.
- De s'affranchir des temps de réponse du réseau : Dans les systèmes distribués les protocoles définissent comment les messages et les données sont échangés. Pour modifier un protocole, il faut changer le code sur toutes les machines du système. Les agents encapsulent les protocoles. Changer de protocole revient à interagir avec un nouvel agent ;
- L'autonomie des composants et l'exécution synchrone/asynchrone : Les terminaux mobiles (PDA, ordinateurs portables, téléphones, ...) ne sont pas toujours connectés au réseau. Les systèmes qui nécessitent des connexions permanentes ne sont pas adaptés dans ce cas. Avec des agents, les terminaux mobiles peuvent se connecter pour vérifier s'ils ont des messages. Un agent peut être envoyé sur le terminal mobile au moment de la connexion et travailler sur ce dernier après avoir interrompu la connexion. Le même agent peut attendre la prochaine connexion pour envoyer des informations sur le réseau signifiant par exemple qu'une tâche a été terminée.

³ Les *systèmes multi-agents* provenant des domaines de la robotique et de l'intelligence artificielle qui proposent de structurer une application en différentes fonctions représentées chacune par un agent. Les différents agents coopèrent pour la résolution des problèmes de l'application.

⁴ Les *agents d'interface* provenant des domaines de l'interface homme-machine et de l'intelligence artificielle qui essaient de simplifier la vie de l'utilisateur en automatisant des tâches réalisées couramment par l'observation des comportements répétitifs ou en surveillant des ressources comme le courrier électronique.

- L'adaptation dynamique du système aux problèmes : Les agents peuvent se répartir eux-mêmes sur des machines du réseau, afin de mieux tenir compte de l'état du système au moment où ils doivent résoudre leurs tâches.
- La gestion des environnements hétérogènes
- La gestion de la robustesse et de la tolérance aux pannes : Si une machine doit s'arrêter, l'agent qui s'y trouve peut changer de machine pour terminer sa tâche en cours.

b. Définition d'un agent mobile

Voici les premières définitions données aux agents mobiles [7]:

Un agent est une entité autonome, réelle ou abstraite, qui est capable d'agir sur elle-même et sur son environnement, qui, dans un univers multi-agent, peut communiquer avec d'autres agents, et dont le comportement est la conséquence de ses observations, de ses connaissances et des interactions avec les autres agents.

Il ressort de cette définition des propriétés clés comme *l'autonomie*, *l'action*, la *perception* et la *communication*. D'autres propriétés peuvent être attribuées aux agents. Nous citons en particulier la *réactivité*, la *rationalité*, *l'engagement* et *l'intention*.

Un agent est un système informatique, *situé* dans un environnement, et qui agit d'une façon *autonome* et *flexible* pour atteindre les objectifs pour lesquels il a été conçu.

Les notions "situé", "autonomie" et "flexible" sont définies comme suit:

- *situé*: l'agent est capable d'agir sur son environnement à partir des entrées sensorielles qu'il reçoit de ce même environnement. Exemples: systèmes de contrôle de processus, systèmes embarqués, etc ;

- *autonome*: l'agent est capable d'agir sans l'intervention d'un tiers (humain ou agent) et contrôle ses propres actions ainsi que son état interne;

- *flexible*: l'agent dans ce cas est:

capable de répondre à temps: l'agent doit être capable de percevoir son environnement et élaborer une réponse dans les temps requis,

proactif: l'agent doit exhiber un comportement proactif et opportuniste, tout en étant capable de prendre l'initiative au "bon" moment;

social: l'agent doit être capable d'interagir avec les autres agents (logiciels et humains) quand la situation l'exige afin de compléter ses tâches ou aider ces agents à accomplir les leurs.

Donc, nous pouvons définir un agent comme un programme informatique qui possède les quatre propriétés suivantes :

Autonomie : Un agent contrôle ses actions et son état interne

Communication : Un agent peut communiquer avec d'autres agents

Réactivité : Un agent perçoit l'environnement dans lequel il se trouve et peut répondre aux changements qui s'y produisent.

Activité : Un agent est capable d'agir sur l'environnement dans lequel il se trouve

c. Particularités des agents mobiles

La migration « forte/faible » : la migration permet le transfert d'un agent d'un site à un autre à travers le réseau. Le problème principal vient du fait que nous ayons à faire à des agents, c'est-à-dire des entités autonomes (pro-actives) qui répondent à des sollicitations externes et internes (leurs propres sollicitations). De ce fait l'agent mobile peut :

- être sollicité de l'extérieur pour faire une migration (sur ordre de l'administrateur de la machine hôte avant un arrêt de la machine) alors qu'il est en train d'effectuer des traitements
- ou la solliciter lui-même alors qu'il est sollicité de l'extérieur.

Deux types de migrations ont été proposés dans les plate-formes existantes : la migration forte permet à un agent de se déplacer quelque soit l'état d'exécution dans lequel il se trouve et de reprendre son exécution après la migration exactement là où elle en était avant ; la migration faible ne fait que de transférer avec l'agent son code et ses données et soit nécessite des moments privilégiés dans le code de l'agent pour pouvoir être lancée (point d'arrêt), soit peut être lancée à tout moment mais avec perte des traitements en cours. La migration forte, bien que beaucoup plus exigeante à implanter que la migration faible, n'en est pas moins indispensable pour toutes les applications pour lesquelles les notions de fiabilité et de tolérance aux pannes sont primordiales [5].

L'accès aux ressources locales : cela pose en fait le problème du piratage des sites visités par les agents mobiles. Ce problème, non spécifique aux agents mobiles, est symétrique pour ces

derniers. En effet un agent mobile est aussi à la merci de se faire pirater par les logiciels qui fonctionnent sur les sites qu'il visite. Cela revient au problème général de la sécurité des machines (lutte contre les virus et autres techniques connues de piratage) avec comme spécificité le fait que les agents mobiles peuvent naturellement se déplacer beaucoup, d'où le problème de leur traçabilité.

La traçabilité : cela peut être **un moyen pour suivre et contrôler les déplacements des agents**. Les agents mobiles peuvent disposer de toute l'autonomie leur permettant de choisir leurs itinéraires (espace) et les moments (temps) pour les explorer. Par exemple il peut être judicieux pour un agent de changer l'ordre des sites qu'il doit visiter à cause d'une indisponibilité passagère d'un lien réseau ; il peut aussi décider de ne pas remettre en question l'ordre de son itinéraire, mais simplement d'attendre un peu que la perturbation du lien réseau s'estompe. On peut aussi avoir des agents dont les missions sont soit exploratoires (robots sur Internet), soit font référence à des buts ou à des besoins, charge aux agents de trouver les bons interlocuteurs pour mener à bien leurs tâches. Bref, même si les agents peuvent être programmés pour donner des informations sur ce qu'ils font, il peut être utile de **disposer de mécanismes de traçabilité qui leur sont externes**, pour garder un aperçu de ce qu'ils font.

La communication entre agents : on peut identifier un grand nombre de moyens de communication entre agents, mais on peut les classer suivant qu'il s'agit d'une communication entre deux agents ou entre un agent et un groupe d'agents (groupe au sens large, c'est-à-dire soit un groupe formellement identifié en tant que tel, soit un ensemble d'agents non formellement identifiés comme par exemple des agents présents à un moment donné dans un lieu donné – concomitance spatio-temporelle).

Les moyens de communications sont multiples (messages, sessions, tableau blanc, rendez-vous, de type collaboratifs, par événements distribués ou locaux, ...).

d. Quelques systèmes d'agents mobiles

La notion d'agent est assez récente dans le domaine des réseaux et des systèmes répartis et engendre une activité de recherche très importante. Le système *Telescript* de *General Magic*, qui est à l'origine du premier système d'agent mobiles, a déclenché l'intérêt des grands opérateurs de télécommunication pour cette nouvelle forme de structuration des communications. Beaucoup de projets de recherche sont en cours pour à la fois approfondir

cette nouvelle forme d'architecture des réseaux et étudier l'impact de cette technologie sur l'existant en matière de communication et de services.

Dans cette section, nous tentons de donner une idée de l'état actuel des connaissances en matière d'agents. Cette tâche est assez difficile pour plusieurs raisons:

- Cette technologie est en plein essor
- Il existe beaucoup d'implémentations différentes de la notion d'agent.
- Il y a eu plusieurs tentatives de normalisation pour prévenir la profusion d'implémentations incompatibles entre elles.
- Il y a plusieurs modèles pour décrire le comportement des agents. Chacun fait l'objet de recherches et de documents qui ne permettent pas d'avoir une vue globale et homogène.
- La maîtrise technologique des agents est un enjeu économique et stratégique très important, ce qui explique la difficulté de trouver une normalisation acceptable. En effet les entreprises qui ont déjà implémenté leur système d'agents ont du mal à se mettre d'accord sur un standard.
- La technologie des agents pose des problèmes théoriques et pratiques, notamment en terme de sécurité, très difficiles et dont l'étude en est encore au stade de la recherche.

Nom	Source	Licence
AgentOWL	http://agentowl.sourceforge.net/	open source
Akira	http://akira-project.org/	open source
LPA Agent Toolkit	http://www.lpa.co.uk/atk.htm	Academic free
MASS	http://mas.cs.umass.edu/research/mass/index.html	open source
SPYSE	http://www.zope.org/Members/drapmeyer/spyse	open source
Tryllian ADK	http://www.tryllian.com/technology/product1.html	open source
WSDL2Agent	http://sas.ilab.sztaki.hu:8080/wsdl2agent/	open source

Systemes d'agents mobiles en 2005

Les plus populaires	Source
Aglets	IBM - http://www.trl.ibm.com/aglets/
Voyager	http://www.recursionsw.com/products/voyager/voyager.asp
D'Agents	http://agent.cs.dartmouth.edu/general/agenttcl.html
JADE	http://sharon.csel.it/projects/jade/
MadKit	http://www.madkit.org

Les systèmes d'agents mobiles les plus populaires

IV. CONCLUSION

L'avènement des grilles permettra de gagner plusieurs ordres de grandeur dans la taille des problèmes résolus en calcul scientifique. Les premiers projets de grilles informatiques ont montrés leur faisabilité avec le développement d'infrastructure matérielle et logicielle et des expériences concluantes à l'échelle de l'Internet et d'un intranet. De nos jours, le domaine industriel nécessite d'avoir recours à des environnements de calcul haut débit tels que des grilles pour simuler des comportements réels ou simplement vérifier une cohérence particulière sur des produits virtuels en cours de définition.

Ce sont aussi des systèmes complexes impliquant la mise en oeuvre cohérente de nombreuses couches logicielles. Globalement, bien que les premières grilles de production soient annoncées, il reste de très nombreuses questions et problèmes à résoudre pour que les grilles soient accessibles par des utilisateurs non spécialistes. Au-delà, les propriétés attendues de performance et de tolérance aux pannes ouvrent un vaste espace d'optimisations. Dans ce contexte, il nous parait clair que les notions de mobilité ont un rôle central à jouer dans les grilles informatiques. L'utilisation des agents mobiles peut permettre d'améliorer les performances des communications. En déportant des calculs, ils favorisent la diminution de la communication en bande passante à travers le réseau. Bien que ces agents posent de nouveaux problèmes de sécurité, ils peuvent permettre des interactions distantes robustes sur des réseaux soumis à des défaillances de canaux de communication ou de machines.