

Avakas : un an de production au Mésocentre Aquitain

Plan

- MCIA
 - MCIA : bref historique
 - Renaissance 2010
 - Ressources
- Avakas
 - La machine
 - Politique d'utilisation
 - Scheduler
 - Files d'attente et priorités
 - Environnement et services
 - Résultats : Utilisateurs
 - Résultats : Production
 - Résultats : Retombées scientifiques
- Conclusion, directions futures

MCIA : bref historique

- 1993 : Modélisation Numérique Intensive (MNI)
 - Fondé par 4 laboratoires de l'Université Bordeaux 1
- 2003 : Mésocentre M3PEC
 - Université Bordeaux 1
- 2010 : Mésocentre MCIA
 - PRES – Université de Bordeaux
 - Financement Conseil Régional + FEDER + établissements

Renaissance 2010

- Nouveau périmètre
 - Universités bordelaises, IPB, CNRS, INRA, INRIA, INSERM, Université de Pau et des Pays de l'Adour, ...
 - PME
- Nouvelle missions
 - Opérer les ressources de calcul pour la communauté scientifique régionale
 - Fournir des services pour la promotion du calcul scientifique
- Nouvelle organisation
 - Comité scientifique – participation ouverte
 - Comité des utilisateurs – participation ouverte
 - Groupes thématiques ouverts : appels d'offre, etc... - participation ouverte
 - Equipe technique – Direction Informatique U Bordeaux 1



Ressources

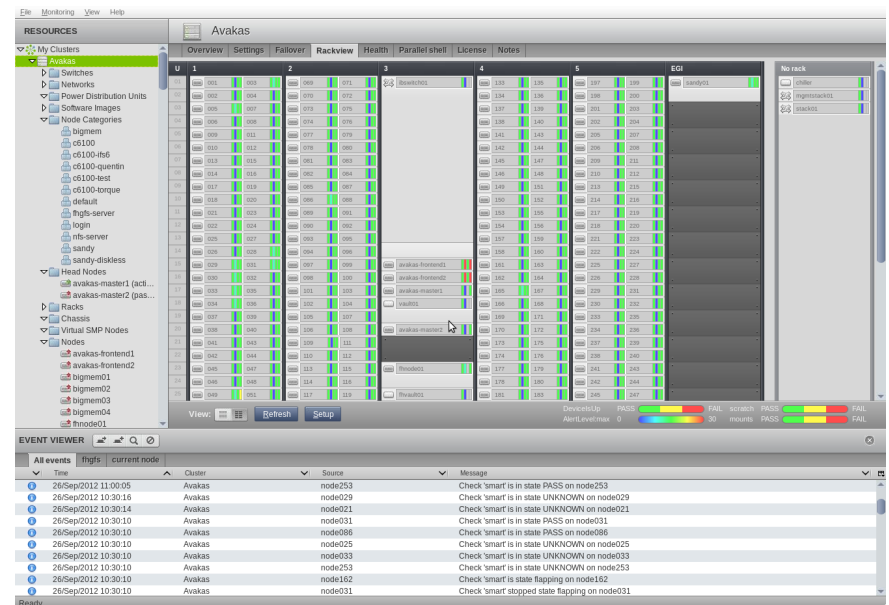
- Avakas : le nouveau cluster HPC
 - En production le 25/11/2011
- M3PEC : le site grille EGI
 - 2 clusters : Bx1 + INRA
 - En production depuis fin 2009
 - Extension INRA : juin 2012
- Personnels
 - 2 ingénieurs permanents U. Bordeaux 1
 - 2 ingénieurs contractuels (3 ans)

Plan

- MCIA
 - MCIA : bref historique
 - Renaissance 2010
 - Ressources
- Avakas
 - La machine
 - Politique d'utilisation
 - Scheduler
 - Files d'attente et priorités
 - Environnement et services
 - Résultats : Utilisateurs
 - Résultats : Production
 - Résultats : Retombées scientifiques
- Conclusion, directions futures

La machine

- La commission d'appel d'offres
 - 18 participants
 - Objectif : une machine HPC de production qui permet aussi des expériences
 - Dialogue compétitif sur 18 mois
 - 4 constructeurs candidats
 - 1 lauréat : Clustervision+DELL+QLogic+Rittal
- Caractéristiques techniques
 - Partition C6100
 - _ 264 nœuds à 12 cœurs (3168 total) cœurs Intel à 3,06 GHz
 - _ 48 Go RAM par nœud
 - Partition Bigmem
 - _ 4 nœuds à 40 cœurs (160 total) cœurs Intel à 2,4 GHz
 - _ 512 Go RAM par nœud
 - _ 4 disques SAS 15000rpm
 - Réseau
 - _ 1 switch Infiniband QDR Qlogic 40 Gb/s 324 ports
 - Système de fichiers parallèle FhGFS 5Go/s agrégés
 - Administration
 - _ Bright Cluster Management
 - Système de refroidissement par eau Rittal



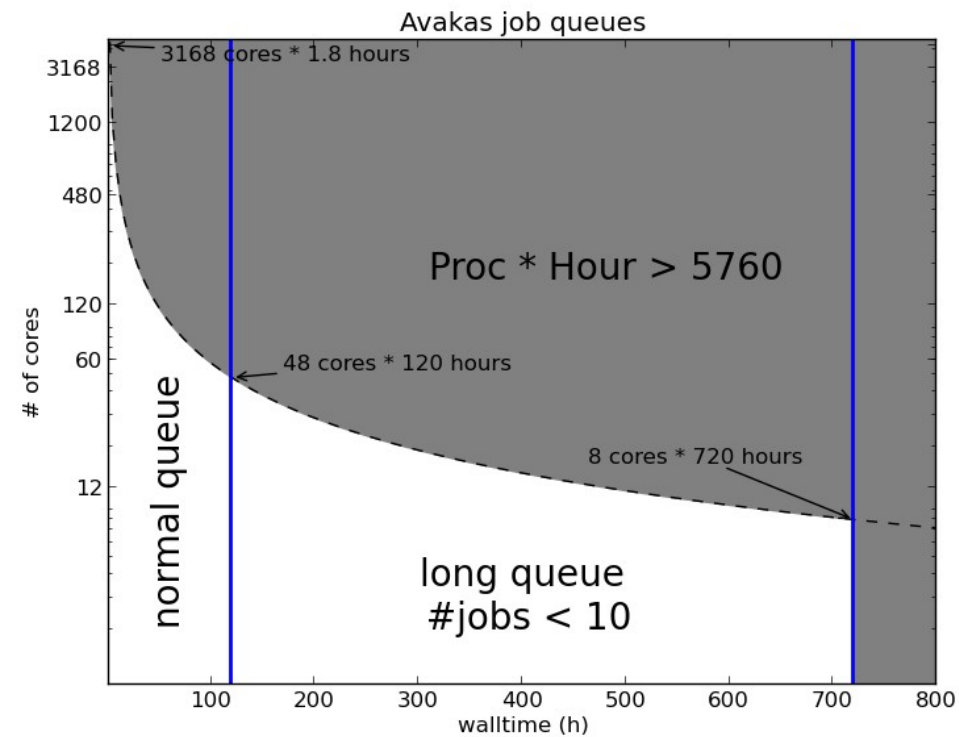
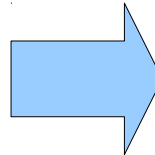
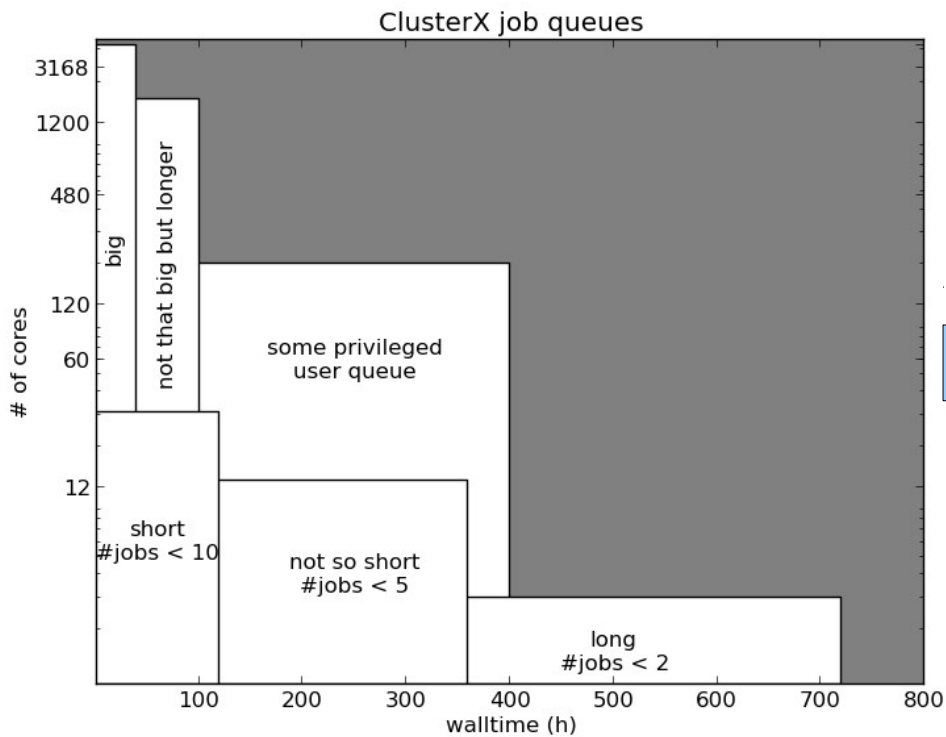
Politique d'utilisation

- Définie avec le Comité des utilisateurs
 - Objectif : restituer le maximum de ressources
- Accès aux ressources
 - Ouvert à tous les chercheurs des établissements du PRES
 - Soumis à une charte d'utilisation
 - Notamment citation correcte du MCIA dans les articles scientifiques
 - Fourniture d'une liste des articles pour le rapport scientifique du Mésocentre
- Répartition des ressources
 - Pas de comité d'attribution
 - Répartition à posteriori
 - Sur un modèle inspiré de la Grille EGI
 - Attirant pour les nouveaux utilisateurs

Scheduler

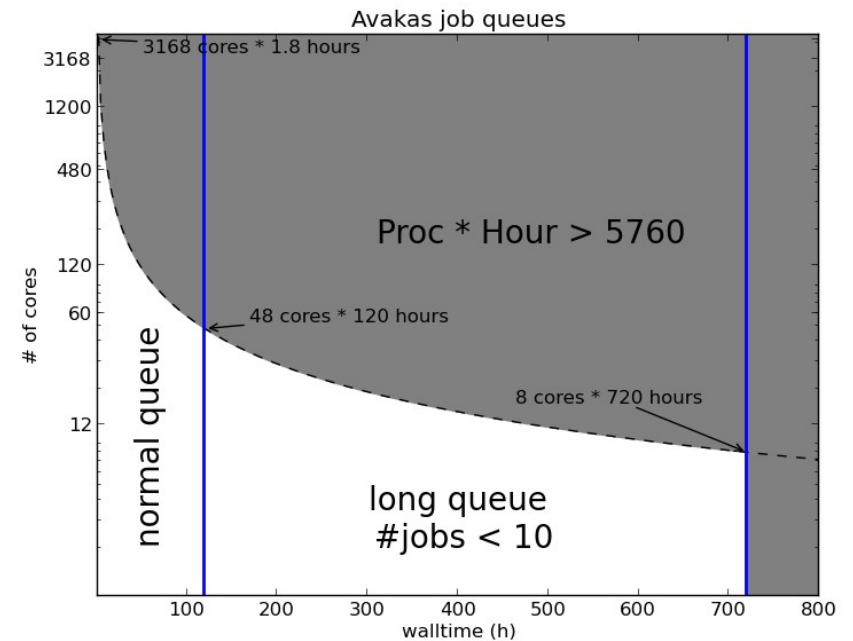
- Torque + MAUI
 - Garantir une séparation raisonnable des jobs
 - Favoriser le roulement des jobs (jobs courts)
- Contraintes
 - Jobs confinés dans des « cpuset » Linux
 - Pas d'accès ssh aux nœuds pour garder le confinement (workaround disponible)

Files d'attente et priorités



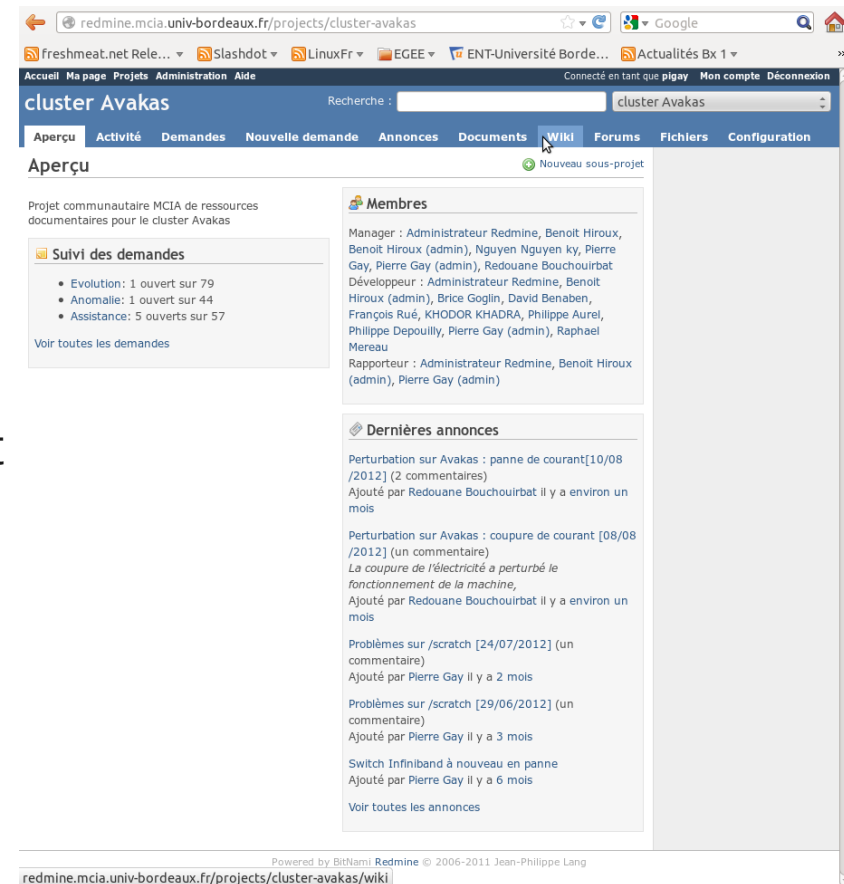
Files d'attente et priorités

- Objectif
 - Incitation au calcul parallèle
 - Sans toutefois l'imposer
 - Lisibilité des règles de scheduling
- Limites aux jobs
 - File d'attente « groupe » : walltime < 120 h
 - File d'attente « longq » : walltime < 720 h, 10 jobs simultanés par utilisateur
 - Proc*Hour < 5760 p*h par job
 - Par utilisateur : MAXPS < 25000 p*h soft (50000 p*h hard)
 - Dérogation possible avec l'accord du Comité des utilisateurs
- Définition des Priorités
 - FairShare par utilisateur et par groupe
 - XFactor : priorité aux jobs courts
 - Backfilling



Environnement et services

- Serveur Redmine communautaire
 - <http://redmine.mcia.univ-bordeaux.fr>
 - Wiki (public)
 - Forums (public)
 - Tickets (authentification)
- Software admins
 - Des utilisateurs qui le demandent installent et maintiennent des logiciels pour la communauté
 - 23 personnes, 64 logiciels contribués
- Services à la communauté
 - Formations : présentation du cluster, programmation parallèle, ...
 - Aide à l'optimisation, la parallélisation
 - Participation à la conception des projets



Résultats : Utilisateurs

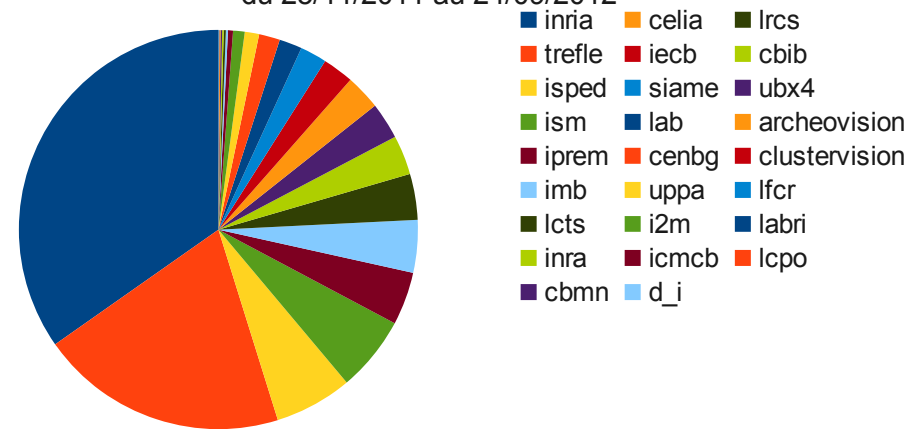
- Inscrits : 210
 - 160 ont effectivement lancé un calcul
 - 111 ont utilisé plus de mille heures de calcul (hCPU)
 - 30 ont fait plus de 100 000 hCPU
- Renouvellements demandés (septembre 2012) : 107
- Certains laboratoires qui n'avaient pas de culture du calcul intensif ont produit jusqu'à 1 million d'heures CPU

Résultats : Production

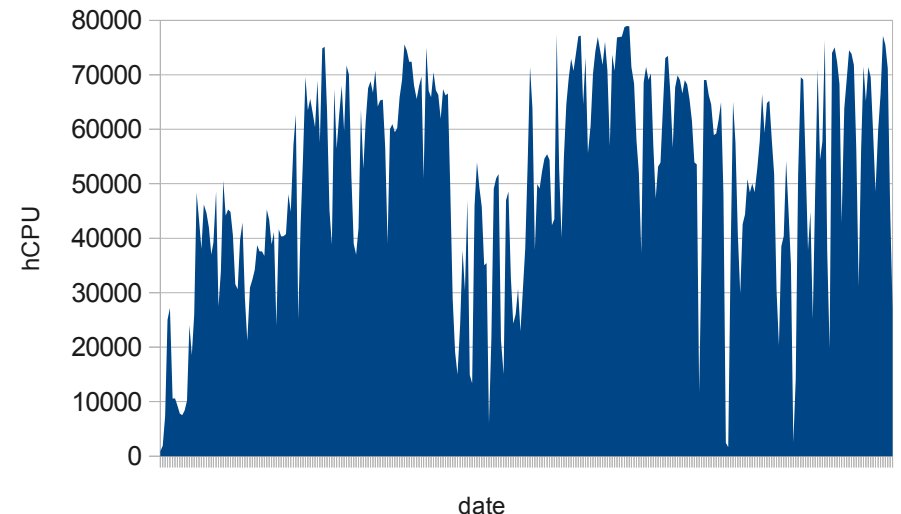
- En hCPU : *procs x walltime consommés*
- Période du 25/11/2011 au 24/09/2012
 - 24,2 millions hCPU possibles
 - 15,6 millions hCPU produites
 - 64,5 % de charge

Répartition heures CPU

du 25/11/2011 au 24/09/2012



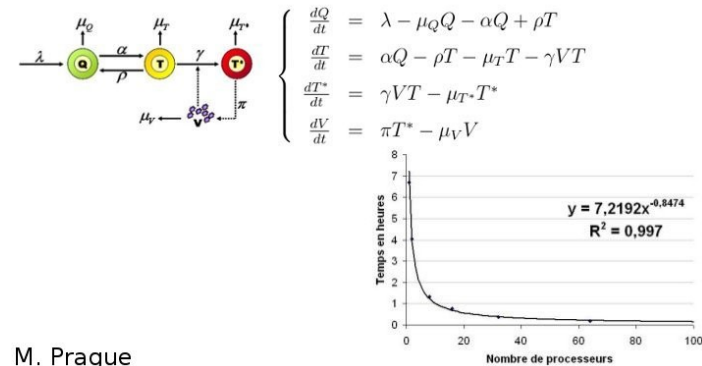
Production Avakas



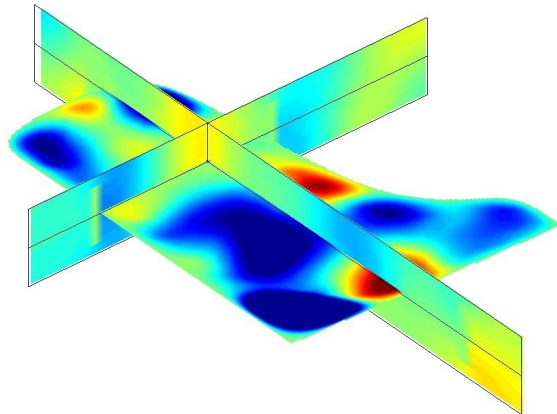
Résultats : Retombées scientifiques

- Contributions obtenues lors de la campagne de renouvellement des comptes : 70 articles, posters, etc...

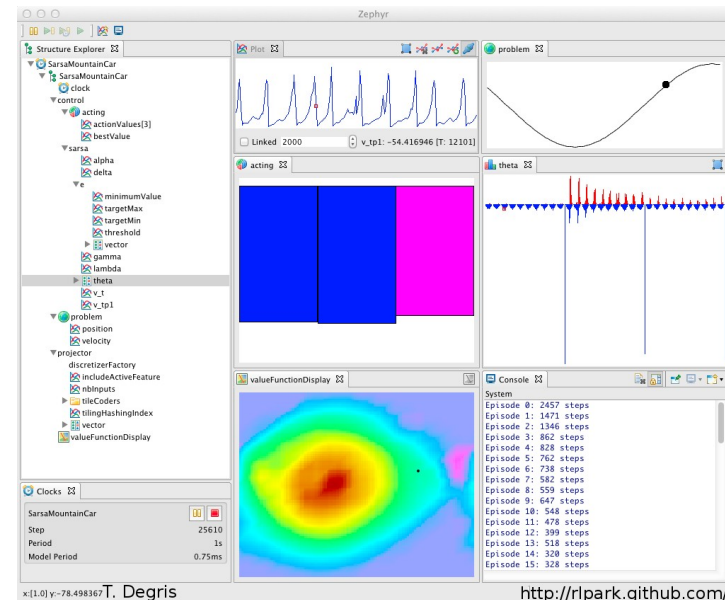
Modèle à cellules T actives¹



M. Prague

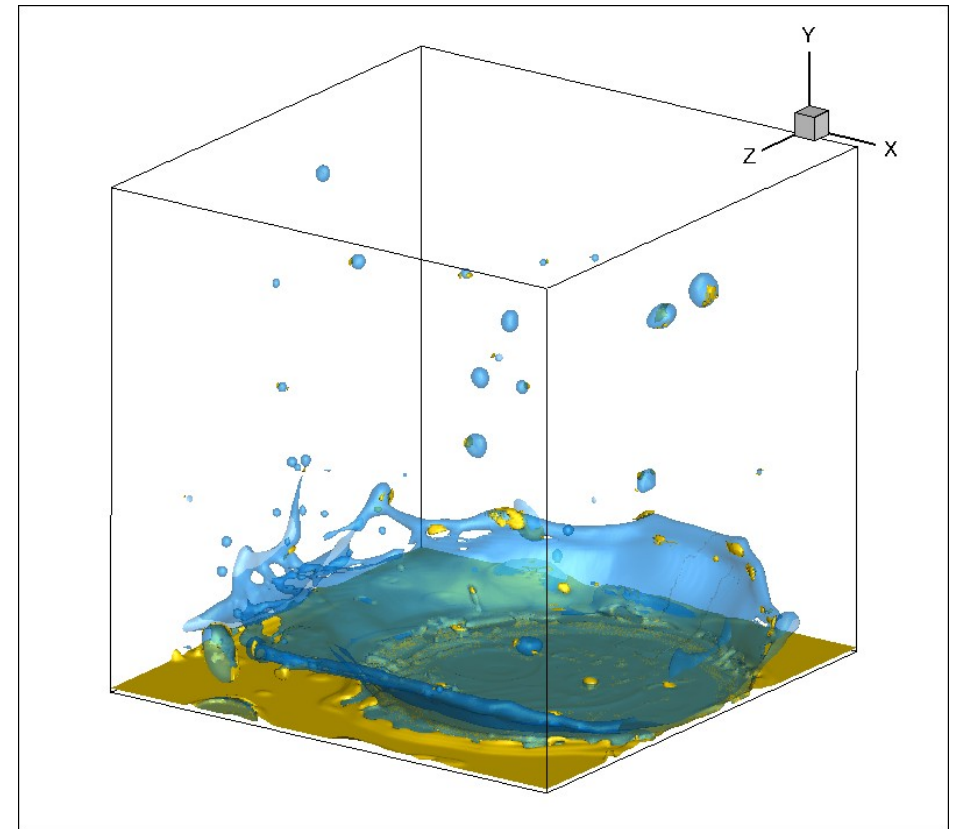
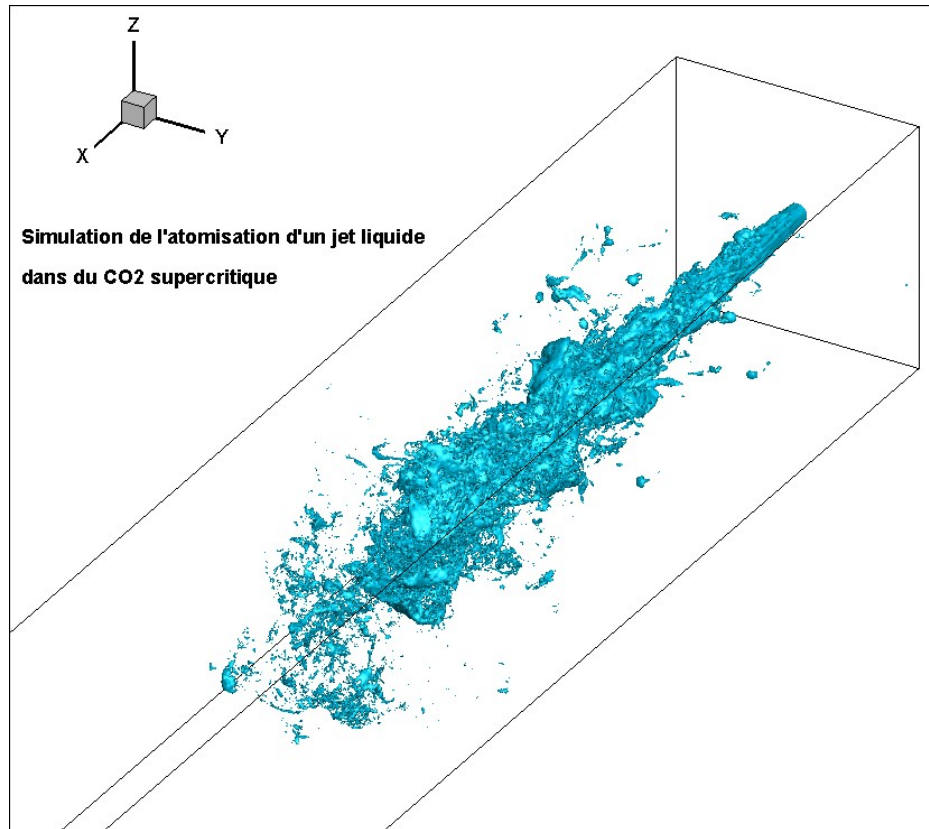


J. Chabassier
<http://modelisation.piano.free.fr/Piano/Welcome.html>



<http://rlpark.github.com/>

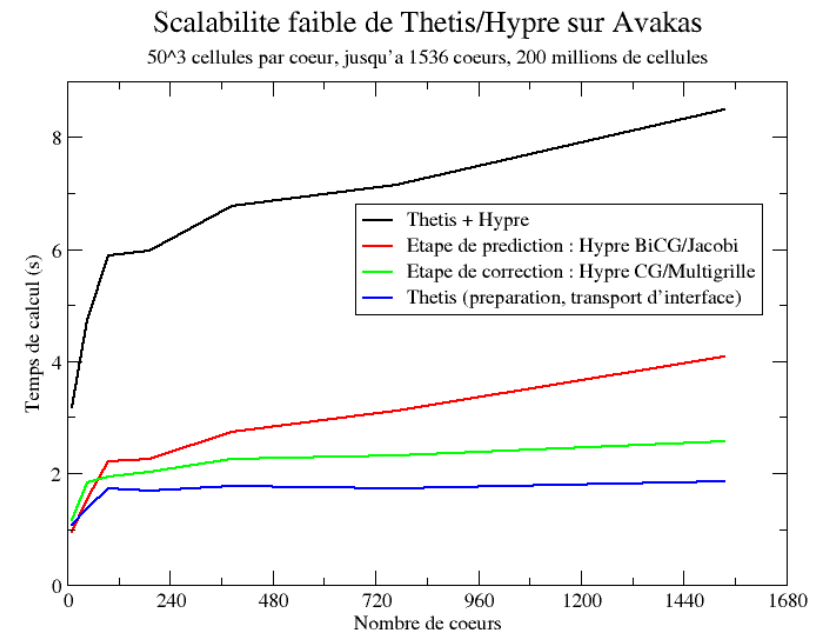
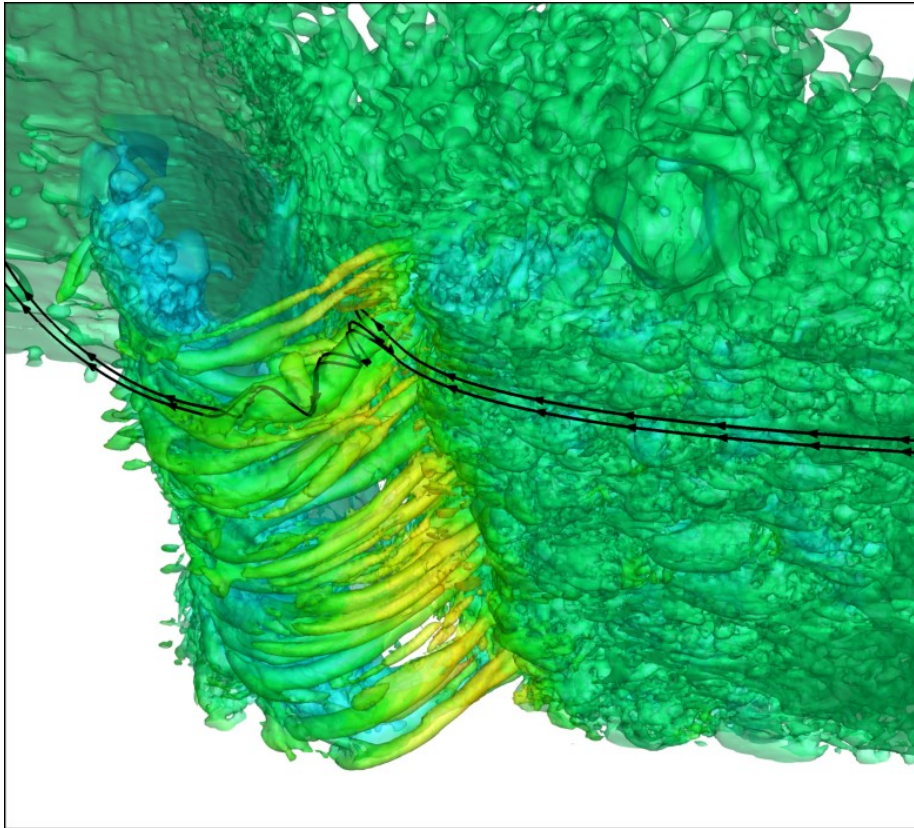
Résultats : Retombées scientifiques



S. Glockner, P Lubin, A. Erriguible, R. Ouchene

<http://thetis.enscbp.fr>

Résultats : Retombées scientifiques



S. Glockner, P Lubin, A. Erriguible, R. Ouchene

<http://thetis.enscbp.fr>

Plan

- MCIA
 - MCIA : bref historique
 - Renaissance 2010
 - Ressources
- Avakas
 - La machine
 - Politique d'utilisation
 - Scheduler
 - Files d'attente et priorités
 - Environnement et services
 - Résultats : Utilisateurs
 - Résultats : Production
 - Résultats : Retombées scientifiques
- Conclusion, directions futures

Conclusion, directions futures

- Enseignements :
 - L'organisation ouverte du Mésocentre mène naturellement à une structuration et un fonctionnement communautaires
 - Les utilisateurs ont besoin de ressources, mais aussi de services et de conseil
 - La prochaine fois : choix d'un système de fichiers parallèle plus ouvert ?
- Futur :
 - Visualisation déportée / GPU : *coming soon...*
 - Poursuite de l'investissement dans le monde de la grille (INRA Bordeaux, DIRAC France Grilles, ...)
 - Cloud Computing ?
 - Quid des données ?

Réactions, Questions ?

- Conseil scientifique :
 - Jean Christophe Soetens :
jsoetens@ism.u-bordeaux1.fr
- Equipe technique :
 - Pierre Gay : pierre.gay@u-bordeaux.fr
- Site web :
 - <http://www.mcia.univ-bordeaux.fr>