

Executive summary

L'IN2P3 et l'IRFU sont à la pointe du progrès dans le traitement des très gros volumes de données depuis de nombreuses années. Poussées par les besoins de la physique des particules, elles ont mis en place les outils et acquis des ressources dont bénéficient toutes les autres disciplines scientifiques, notamment à l'IRFU et à l'IN2P3 : astroparticules, physique nucléaire et hadronique, recherches pluridisciplinaires.

Au cours des dix dernières années, la physique des particules a joué un rôle majeur dans le développement de la technologie des grilles informatiques. Le pari pris en 2001 de miser sur la grille pour l'analyse des données du LHC est une totale réussite. En 2011, cette analyse mobilise environ 250.000 processeurs dans le monde sur la LHC Computing Grid tandis que 150 PetaOctets sont nécessaires au stockage des données. L'augmentation progressive de la luminosité et l'upgrade des détecteurs va générer une croissance continue du besoin en ressources informatiques dans les 10 prochaines années. L'évolution du modèle de calcul et notamment l'accroissement des échanges de données entre Tiers 2 requièrent une amélioration des performances du réseau entre ceux-ci. La France fournit aujourd'hui 10% des ressources de la grille du LHC et l'objectif est de maintenir ce niveau de contribution mais les deux dernières années (2010, 2011) ont été particulièrement décevantes en termes de financement par le ministère de la recherche et ce niveau de participation ne pourra être maintenu à un tel niveau de financement. En plus des besoins du LHC, le démarrage de SuperB va générer des volumes de données significatifs à l'échelle de quelques PBytes par an tandis que les inconnues qui subsistent sur le calendrier de l'ILC conditionné par les résultats de physique au LHC rendent toute extrapolation impossible sur l'évolution des besoins.

Dans le même temps, la physique des astroparticules va connaître un changement d'échelles dans le volume des données à traiter sur une nouvelle génération d'instruments. Les équipes et les centres de ressources français (notamment le CC-IN2P3) sont aujourd'hui des acteurs majeurs du traitement des données de nombreuses expériences (AMS, ANTARES, Auger, Fermi, HESS, VIRGO) et se positionnent (notamment le Centre F. Arago à l'APC et le LAPP) sur les instruments de l'avenir les plus ambitieux (LISA, EUCLID, LSST, CTA, ...).

En astrophysique, l'accès à une grande puissance de calcul permet le raffinement des modèles physiques et l'augmentation de la résolution numérique, qui est un enjeu particulièrement important pour de nombreux champs de recherche. Tout ceci a un impact immédiat sur la compréhension des phénomènes (astro)physiques. Les ressources de calcul utilisées sont à la fois les moyens de calcul intensif locaux (serveurs, grappes de serveurs), régionaux (méso-centres) et centralisés (centres nationaux et internationaux). Pour certaines thématiques, il est nécessaire de pouvoir avoir accès à un grand nombre de processeurs (plusieurs centaines à quelques milliers) sur des durées pouvant parfois atteindre la semaine voire le mois.

Dans ce contexte, les moyens en méso-équipement pour le calcul HPC (Tiers 2) doivent être maintenus et pérennisés sur le long terme à l'IRFU et à l'IN2P3 au niveau de 10% des centres nationaux (GENCI) ou 1% des centres européens (PRACE). Ces moyens de calculs Tiers 2 sont indispensables à la formation des jeunes collaborateurs, au développement et à la validation des codes de calculs parallèles HPC développés depuis

plus de 10 ans et la préparation des simulations pour les centres Tiers 0 et 1. Complémentaires aux serveurs de calculs, des moyens puissants pour la visualisation des jeux de données 4-D (3 d'espace et 1 de temps) sont nécessaires pour l'analyse détaillée des simulations HPC.

La physique théorique par l'étude de QCD sur réseau est aujourd'hui un client majeur des grands supercalculateurs nationaux. La compréhension des phénomènes hadroniques pour l'étude de la violation de CP au LHC mais aussi sur les nouvelles machines SuperB et Belle II requiert des calculs toujours plus précis pour lesquels une machine pétaflopique dédiée mobilise la communauté.

A l'heure où nous écrivons ce rapport, le cloud computing apporte une révolution dans le monde du calcul intensif et du traitement de grandes masses de données. En effet, pour la première fois, les industriels ne se contentent pas de vendre des équipements pour le calcul intensif ou à haut débit mais ils développent des offres concurrentes à celles des centres de calculs académiques. Le business modèle du cloud est beaucoup plus attrayant que celui de la grille pour les industriels du secteur et la pression va augmenter considérablement pour une externalisation des services informatiques afin d'en réduire les coûts.

Une telle externalisation ne va pas sans risques : perte de contrôle sur l'accès aux données, risques de dérapage des coûts, dépendances techniques, failles de sécurité... Ces risques doivent être évalués et un juste compromis entre externalisation et maintien d'une expertise en interne doit être trouvé pour préserver au sein de l'IRFU et des laboratoires de l'IN2P3 un réseau d'experts pour l'administration des systèmes, le développement et le déploiement d'applications scientifiques. Les grilles ont permis de structurer et de resserrer ce réseau humain d'experts dont le CC-IN2P3 avec ses ingénieurs est le cœur technique et dont la grille du LHC est la colonne vertébrale. Le défi des prochaines années se situe à plusieurs niveaux :

- offrir aux utilisateurs des grilles de production une qualité de service stable
- développer une infrastructure de clouds académiques adossée aux clouds publics qui se substitue progressivement et de façon transparente à la grille

Le CC-IN2P3 a un rôle essentiel à jouer pour relever ce défi. Ses ingénieurs sont les leaders techniques incontestables de l'infrastructure nationale de traitement massif des données du LHC, quelle que soit la technologie sous-jacente. En complément de l'expertise technique, les nouveaux locaux inaugurés en 2011 lui permettront jusqu'en 2020 de faire face aux besoins croissants du LHC mais aussi de SuperB et des grands instruments pour l'astroparticules.

La sécurité des services informatiques demeure un souci permanent et une priorité. Par leur caractère distribué et ouvert, les grilles ont apporté de nouvelles contraintes. Il en sera de même pour le cloud computing avec notamment tous les problèmes posés par la virtualisation.

Enfin, la préservation des données et la problématique de l'accès à moyen et long terme doivent être considérées comme des nouveaux paradigmes et prises en compte de manière durable dans les programmes scientifiques.