

Séminaire de prospective scientifique 2019 du CNES : le Centre François Arago (FACe), un centre d'exploration des technologies pour le traitement de données spatiales

Etat des lieux :

1. Présentation du FACe

Le Centre François Arago (FACe) est le centre de traitement des données spatiales de l'APC. Ce centre a vu le jour en octobre 2010 et vise à fournir l'infrastructure et l'expertise nécessaires pour répondre aux appels d'offre des agences spatiales sur le traitement des données (segment sol). Après huit ans d'existence, le FACe a un plan de charge important qui s'étend sur les prochaines décennies et au-delà. Les infrastructures et les compétences techniques déjà rassemblées au FACe ont permis d'établir ce centre comme un élément essentiel de l'organisation des missions spatiales ou d'expériences pour l'astrophysique et/ou les astroparticules de grande envergure à l'APC et au-delà :

- le centre a accompagné avec succès les équipes de **LisaPathfinder** dans les exercices préparatoires de la mission ;
- ses compétences dans l'utilisation de la virtualisation lui confèrent une position reconnue et stratégique au sein de la communauté de **France Grille** et des réseaux métiers du **CNRS/IN2P3** ;
- son expertise a permis la conception et l'exploitation d'une infrastructure logicielle de développement collaboratif pour l'ensemble de la mission **Euclid**.

Compte tenu de ces résultats, le FACe a été retenu pour héberger le futur Distributed Data Processing Center (DDPC) du grand observatoire d'ondes gravitationnelles **LISA** dont le lancement est prévu en 2034.

Le rôle du FACe a été jusqu'à présent multiple : un centre d'expertise en traitement de données et des méthodes associées pour les missions spatiales, un centre d'accueil des équipes scientifiques pendant les missions et une structure d'interface entre la communauté utilisatrice des données et les ressources nationales comme le centre de calcul de l'IN2P3 (CC-IN2P3) et du CEA. Les collaborations et partenariats entre le CNES et le FACe se sont traduites à travers de nombreuses occasions (projets spatiaux, R&D, prospection...) et c'est dans ce contexte qu'est présentée cette proposition sur de nouveaux outils pour le traitement de données spatiales.

2. Un ambitieux instrument scientifique interdisciplinaire

Le financement du Centre François Arago est assuré par l'IN2P3, l'Université Paris Diderot et les projets spatiaux supportés par le CNES. Fin 2016, il a été décidé de déménager le FACe, qui était installé dans le bâtiment Biopark, au bâtiment Condorcet de l'Université Paris Diderot.

Dans ce contexte transitoire, le FAcE s'est redéfini comme un centre d'expertise au service des projets de l'APC, spatiaux ou au sol. Il a été reconnu comme une **plateforme de recherche de l'IN2P3** et sera labélisé comme telle à partir de 2019. Cette labélisation permettra au centre une ouverture vers l'extérieur afin d'attirer différentes communautés d'utilisateurs.

De plus, la continuité du FAcE a été assurée grâce à l'acceptation en 2017 du projet **DANTE** (multi Data ANalysis ans compuTing Environment for science). Ce projet DANTE est une réponse à l'appel à proposition Sésame de la région Île-de-France en partenariat avec l'IPGP. Ce projet s'inscrit dans la perspective de la future Université de Paris et vise à mettre en œuvre un ambitieux instrument scientifique interdisciplinaire. Sur la période 2019 - 2024, DANTE permettra en effet de fournir :

- une offre commune et fédérée d'infrastructures et de compétences entre l'APC et l'IPGP ;
- un pôle d'expertises et de connaissances multidisciplinaires dirigé par la recherche ;
- une continuité avec le LabEx UnivEarthS (synergie entre les Sciences de l'Univers et les sciences de la Terre) en terme de recherche et de formation ;
- une ouverture claire vers l'Europe et l'international et un lien privilégié avec le spatial.

3. Un centre de R&D sur les technologies des segments sols

L'étude, la mise en place et l'administration de la plateforme de développement collaborative CODEEN (Collaborative DEveloppement ENvironnement) de la mission Euclid a permis d'acquérir une expertise sur l'intégration et le déploiement continu d'applications en production. Cette plateforme CODEEN, dont la gestion a été réalisée en interaction forte avec le CNES et était hébergée au FAcE, a été migrée en 2018 sur le cloud du CC-IN2P3. Cette expertise a été mutualisée en approche multi-projets afin de préparer la phase 0 de la mission LISA.

La préparation de la phase 0 de LISA a donné lieu à une collaboration étroite entre le CNES et les membres du FAcE sur les technologies de conteneurs et la thématique du débordement de calcul d'un cluster de calcul vers un cloud [6]. Le démarrage de la phase A de LISA en 2018 a permis de continuer ce travail [1, 3, 4] en utilisant les solutions de conteneurs pour les services et le calcul.

La participation du FAcE au projet ComputeOps (conteneurs pour le calcul intensif) de R&D transverse de l'IN2P3 démarré en 2018 (durée de 3 ans) a permis de se pencher sur la convergence du calcul intensif et du calcul gouverné par les données (HPDA – High Performance Data Analysis). Dans le cadre de cette activité transverse, une étude de la problématique de portabilité et de répétabilité de l'exécution d'applications conteneurisées sur des architectures de dernières générations (GPU, FPGA) a été commencée. Une application pilote de type Deep Learning utilisant la librairie TensorFlow a été sélectionnée afin de démontrer l'utilisabilité de ce type d'algorithme.

Prospective :

1. Un centre de traitement de données français

Le Centre François Arago participe pleinement au contexte du spatial français. A travers son expérience dans le segment sol de missions spatiales (Euclid, SVOM, LISAPathfinder, LISA) mais aussi d'instruments au sol (LSST, CTA), il a pu devenir un centre d'expertise sur cette thématique. Il a multiplié ses collaborations avec l'IPGP : à travers le LabEx UnivEarthS (projet sur la valorisation des données), le projet DANTE et la future Université de Paris au sein de laquelle le FACe prend part à la discussion de la création d'un pôle spatial de grande ampleur avec une ambition bien différente de l'action structurante « campus spatial » de l'Université Paris Diderot. Ainsi le FACe s'inscrit dans la synergie entre les Sciences de l'Univers et les sciences de la Terre. Il a su acquérir une expertise sur les technologies de virtualisation (cloud, conteneur) [5, 7, 8] qui sont des solutions techniques très utilisées dans les segments sols modernes. Son exploration du domaine du Big Data (Hadoop, Spark) pour les projets LSST et Euclid lui a permis de se préparer à la gestion des futurs grands volumes de données. Le FACe semble donc un lieu privilégié pour l'étude, en vue d'une adoption future, du paradigme informatique et scientifique que représente l'Intelligence Artificielle. Ce travail a déjà commencé dans le cadre du LISA Data Challenge. En effet, ce challenge scientifique vise à fournir des données de simulation afin d'entraîner les codes d'analyse de données en vue de l'écriture de la future chaîne de traitement de données.

2. Un centre ouvert à l'Europe

Dans le paysage international, une collaboration entre le FACe et l'ISDC (Integral Science Data Center) de Genève a commencé dans le cadre d'un observatoire multi-messagers (« MMO – Multi-Messenger Observatory ») pour l'exploration de l'Univers violent [2]. Le projet vise à mettre en place une plateforme de traitements de données en ligne utilisant les dernières technologies de déploiement de services web (cloud, orchestration de conteneurs). Ce projet, ayant pour but de réutiliser les chaînes de traitement de données développés dans les segments sols des projets spatiaux, s'inclut parfaitement dans la prospective du CNES.

Conclusion :

Fort de son expertise, le FACe pourrait ainsi apporter un retour d'expérience au groupe de travail sur le numérique et les données et proposer dans cette thématique son modèle de fonctionnement comme plateforme de segment sol multi-projets. Dans le futur, un (ou des) centre(s) de traitement des données comme le FACe seront indispensables pour la continuation des travaux sur les segments sol des missions spatiales de l'horizon 2020 et plus lointain que sont Euclid, SVOM, LISA. Et cela devra se faire dans le nouveau contexte de l'approche multi-messager. Les méthodes logicielles de pointe comme le Markov Chain Monte-Carlo ou le Machine/Deep Learning dont l'utilisation semble très prometteuse, seront intégrées et améliorées. Des infrastructures matérielles hétérogènes comme les GPU, les FPGA ou la virtualisation seront mises au point continuellement. La convergence des méthodes de calcul intensif et de calcul gouverné par les données verra le jour et contraindra les nouveaux systèmes. Ainsi le

FACe devra continuer à offrir un centre d'exploration des technologies pour le traitement de données spatiales.

Publications et présentations :

[1] C. Cavet, A. Petiteau, M. Le Jeune, **Prototyping for the Distributed Data Processing Center of LISA**, BiDS'19 (2019), In prep.

[2] G. Allen et al., **Multi-Messenger Astrophysics: Harnessing the Data Revolution**, arXiv:1807.04780 (2018)

[3] C. Cavet, **Retour d'expérience sur les conteneurs pour la mission LISA**, Les technologies de conteneurisation dans le cloud et ailleurs, COMET-SIL, ISAE Supaéro, Toulouse, 15 Mars 2018

[4] C. Cavet, A. Petiteau, M. Le Jeune, E. Plagnol, E. Marin-Martholaz, J-B. Bayle, **A proto-Data Processing Center for LISA**, 11th International LISA Symposium, Journal of Physics : Conference Series, Volume 840, conference 1 (2017)
<http://iopscience.iop.org/article/10.1088/1742-6596/840/1/012045>

[5] C. Cavet, **Docker for space missions**, Journées nationales du Développement Logiciel, Marseille, 4 - 7 Juillet 2017
http://devlog.cnrs.fr/_media/jdev2017/poster_jdev2017_dockerspatial_cecile_cavet.pdf?id=jdev2017%3Aposters&cache=cache

[6] M. Poncet, T. Faure, C. Cavet, A. Petiteau, P.-M. Brunet, E. Keryell-Even, S. Gadioux, M. Burgaud, **Enabling collaboration between space agencies using private and cloud based clusters**, BiDS'16 (2016)

[7] C. Cavet, **Review on distributed computing**, Workshop distributed computing in astrophysics, FACe, APC, Paris, 10 - 11 Décembre 2015.
<https://indico.in2p3.fr/event/12042/contribution/1/material/slides/0.pdf>

[8] V. Beckmann, **Flexible Data Processing Solutions for Space Missions**, SCIOPS, ESAC, Madrid, Espagne, 10 - 13 Septembre 2013.
http://www.rssd.esa.int/SYS/CONF2013/include/SCIOPS2013/docs/presentations/20130912-0930-Beckmann_Flexible_SDC_SciOps2013.pdf