

2023

Rapport

Annuel



Sommaire

- 01 **Le Centre de Calcul de l'IN2P3**
- 02 **Le CC-IN2P3 en chiffres**
- 03 **L'organisation au coeur des services**
- 04 **Accompagnement des utilisateurs**
- 05 **Un savoir-faire adapté aux expériences scientifiques**
- 06 **Bilan de Gaz à Effet de Serre (BGES)**
- 07 **Patrimoine informatique**





Centre de Calcul de l'IN2P3

www.cc.in2p3.fr

Le Centre de Calcul de l'IN2P3

Le Centre de Calcul de l'IN2P3 (CC-IN2P3), est une Unité d'Appui et de Recherche du CNRS dont la mission première est la conception et l'opération de services numériques adressant les besoins des expériences des domaines de la physique nucléaire, de la physique des particules et astroparticules.

Il fournit ainsi aux expériences de l'Institut National de Physique Nucléaire et de Physique des Particules (IN2P3), des ressources informatiques nécessaires au transport, au stockage, à la mise à disposition et au traitement des énormes quantités de données qu'elles génèrent. Il propose également un ensemble de services communs collaboratifs, à destination des personnels de l'Institut et des collaborations scientifiques dans lesquelles l'IN2P3 est partie prenante facilitant ainsi les interactions et les échanges : messagerie électronique et instantanée, plateformes diverses de gestion électronique de documents, de gestion de projets, de gestion de développements logiciels, d'hébergement web, de virtualisation etc.

Parmi les expériences utilisatrices des moyens informatiques du CC-IN2P3 se retrouvent celles liées à l'accélérateur LHC du CERN (ATLAS, CMS, ALICE, LHCb), à l'accélérateur GANIL/Spiral 2, à l'observatoire Rubin (LSST) ou encore à l'équipement KM3NeT. Point commun : elles figurent toutes sur la feuille de route nationale des infrastructures de recherche du Ministère de l'Enseignement Supérieur et de la Recherche ou sur la feuille de route européenne ESFRI.

L'exploitation d'une infrastructure informatique conçue pour relever les défis émanant de ces communautés de recherche nécessite une adaptation et une évolution permanente. Ce changement constant doit tenir compte des prochains défis qui découleront du déluge de données du prochain projet LHC à haute luminosité au CERN, du nouvel accélérateur de particules SPIRAL 2 au GANIL, du détecteur BELLE II au Japon et de plusieurs projets de physique des astroparticules ou de cosmologie comme le futur télescope LSST, le satellite européen EUCLID.

Le CC-IN2P3 se doit ainsi d'améliorer constamment ses services pour rester un outil essentiel de la recherche scientifique.

Les pages qui suivent ne constituent pas une description exhaustive de l'ensemble des travaux réalisés au cours de ces deux dernières années, mais mettent plutôt en avant quelques évolutions significatives.

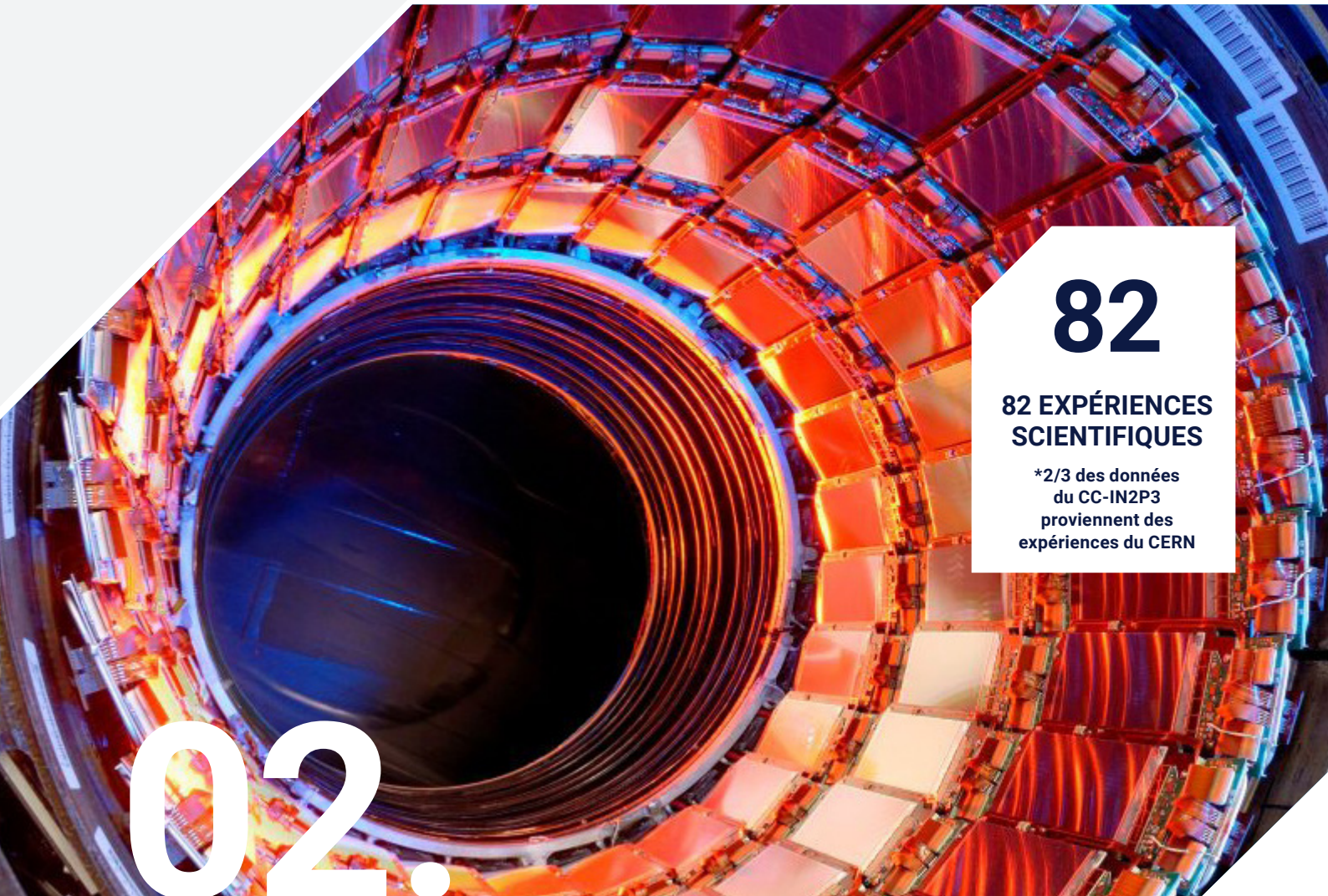




Pierre-Etienne Macchi / Directeur du Centre de Caclul

01.

Le CC-IN2P3 en chiffres //



82

**82 EXPÉRIENCES
SCIENTIFIQUES**

*2/3 des données
du CC-IN2P3
proviennent des
expériences du CERN

02.

55 000

Nombre de programmes que la plateforme de calcul peut exécuter simultanément.

14 000

Nombre de mouvements de bandes magnétiques chaque jour dans les bandothèques du CC-IN2P3.

1 700

Nombre en m2 de la superficie totale des 2 salles serveurs (locaux techniques).

1 000

Nombre de serveurs qui composent la plateforme de calcul et de traitements des données.

482

482 millions d'heures de calcul disponibles par an.

400

Nombre de To de données échangées par le réseau chaque jour entre le CC-IN2P3 et ses partenaires.

237

Nombre de Po de données stockées sur les installations du CC-IN2P3 fin 2023.

5 500

Nombre en m2 du bâtiment universitaire que le système de récupération de chaleur produite par le CC-IN2p3 permet de chauffer partiellement et à moindre coût.

80 //

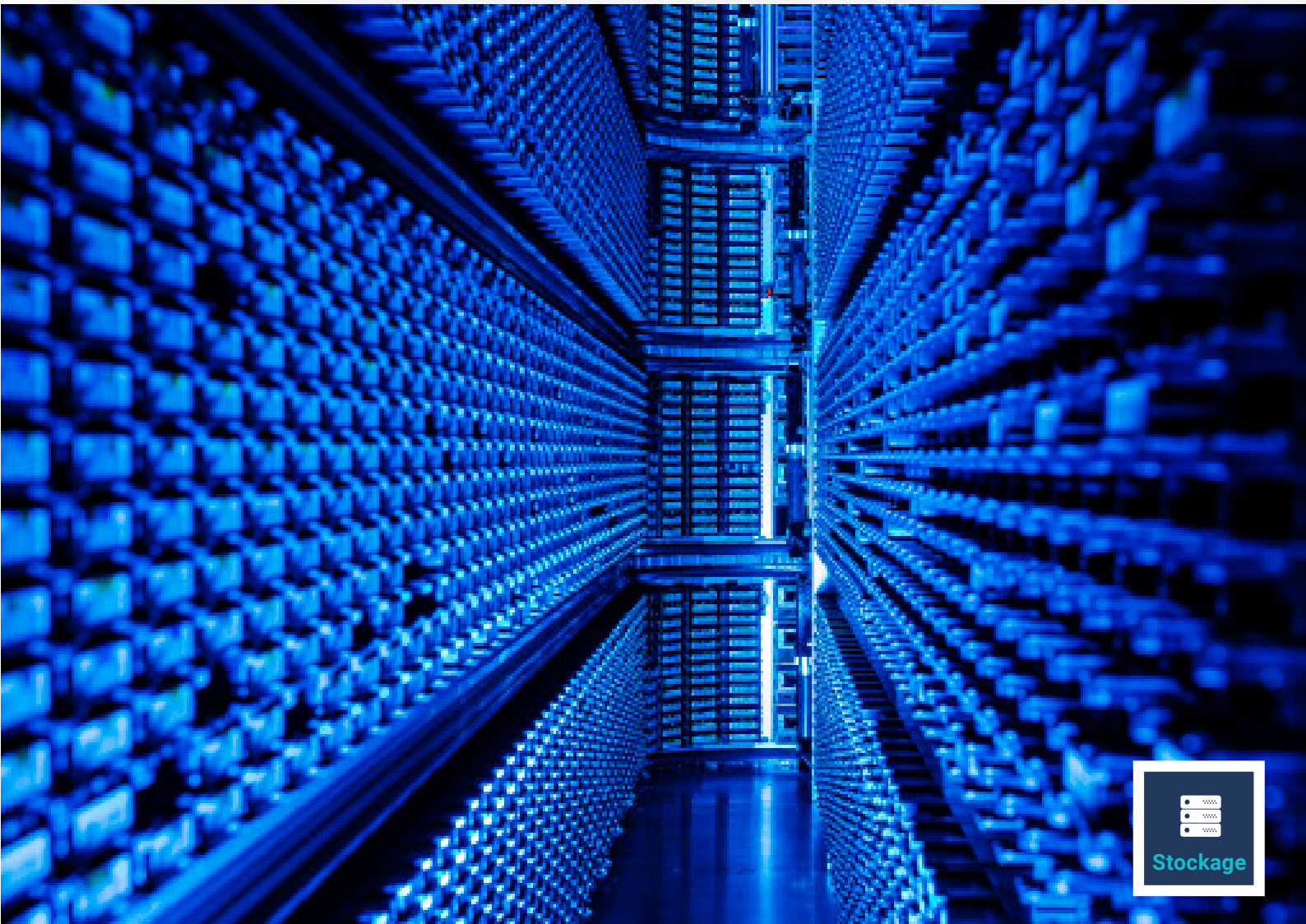
80 agents : c'est l'effectif comprenant les personnels techniques, administratifs et chercheurs impliqués dans la gestion de l'infrastructure et des services informatiques du CC-IN2P3.

365 //

365 jours / an : c'est l'engagement de disponibilité du CC-IN2P3 envers ses communautés d'utilisateurs, pour les services qu'il propose.

Organisation au coeur des services

Stockage - Calcul - Réseau



Stockage

Au cours des quatre dernières années, la quantité de données hébergées sur les divers systèmes de stockage du CC-IN2P3 a augmenté de 25%, passant d'un total de 190 Po en janvier 2021 à 237 Po fin 2023. Cette augmentation a un impact important sur les systèmes de stockage sur disques, principalement en termes d'espace physique occupé (nombre de serveurs et de racks), de consommation d'énergie, de budget et de variation des performances. Cette situation a conduit à développer quatre axes d'évolution afin de garantir la pérennité et la qualité du service.

Augmenter l'utilisation du stockage de masse sur bandes par rapport au stockage sur disques.

S'appuyer davantage sur le stockage sur bandes plus vertueux sur le plan énergétique et financièrement avantageux, implique le déploiement d'une stratégie efficace pour la lecture et la recopie des données de la bande au disque. Pour cela, le développement d'un ordonnanceur de requêtes a permis de séquencer les demandes de lectures des données sur bandes plus efficacement. Jusqu'à 7000 requêtes de montage de bandes ont été traitées par jour pour un volume quotidien de 500 To de données transférées de bandes à disques. Succédant à près de 3 décennies d'utilisation de bibliothèques StorageTek/Oracle, le CC-IN2P3 utilise depuis 2020 des bibliothèques de bandes SpectraLogic TFinity. L'installation de la deuxième bibliothèque SpectraLogic en 2022 a permis d'achever le déplacement des dernières données contenues dans l'infrastructure Oracle. Cette recopie de 80 Po de données a été réalisée en 2 ans et demi tout en conservant parallèlement les activités de production.

Promouvoir l'utilisation de systèmes de fichiers à coût réduit.

L'utilisation des systèmes de fichiers sur les disques a du évoluer. L'espace de travail fourni pour les données à courte durée de vie était principalement basé sur des solutions commerciales IBM Spectrum Scale avant 2019 et Dell EMC Isilon depuis 2018. Pour des raisons principalement financières, le CC-IN2P3 s'est progressivement tourné depuis 2021 vers le déploiement de plateformes CephFS présentant l'intérêt évident d'une meilleure maîtrise des coûts. La technologie CephFS est amenée à remplacer progressivement Dell EMC Isilon à l'horizon 2026. IBM Spectrum Scale a quant à lui été définitivement arrêté à l'été 2023.

Améliorer du cycle de vie des données.

La mise en place progressive des Plans de Gestion des Données (PGD) que doit fournir chaque expérience est maintenant systématique après des années d'ajustement. Couplé à un suivi précis des données stockées sur les systèmes du CC-IN2P3, les PGD permettent de donner un cadre pour la gestion des données stockées dans le temps et d'effectuer la curation et le nettoyage des données devenues inutiles.

Améliorer les performances des services de stockage.

Travailler sur les performances du service de sauvegarde a permis d'améliorer le temps de restauration des données. A la vue des volumétries que certains projets engendrent, le temps de restauration en cas de sinistre partiel ou total peut vite devenir pénalisant. La copie primaire de la sauvegarde qui est sollicitée lors d'une restauration, a été basculée de la bande magnétique au disque. Si ceci peut sembler en contradiction avec la volonté d'utiliser plus de bandes magnétiques, les « faibles » volumétries en jeu (quelques Po) dans le cas précis de la sauvegarde, autorisent cet écart afin de conserver un service viable.

03.

Calcul

Les plateformes de traitement de données CC-IN2P3 sont composées d'un millier de serveurs disponibles 24h/24 et 7j/7, permettant l'exécution d'environ 55000 programmes de calcul simultanément.

L'équipe Calcul en charge de ces plateformes a été créée fin 2019 avec une mission claire : déployer et mettre en œuvre des solutions logicielles essentielles pour le traitement des données. Cette mission comprend des activités de sélection de développement et de déploiement de solutions logicielles aux services d'exploitation et la planification stratégique de leur évolution continue.

Des transformations importantes ont été menées ces dernières années, conduisant à la mise en place de deux plateformes de calcul distinctes. Ce vaste chantier a été guidé par plusieurs objectifs : résoudre les défis opérationnels croissants, améliorer l'intégration du CC-IN2P3 dans le réseau informatique international du LHC et contrôler puis réduire les coûts associés à l'utilisation de logiciels commerciaux. En priorité, Univa Grid Engine, dont le modèle de coût dépendait du nombre grandissant de processeurs.

Depuis 2020, la plateforme intégrée à l'infrastructure de calcul grille WLCG, utilise l'ordonnanceur HTCondor. Avec 70% de la capacité de calcul totale du CC-IN2P3, cette plateforme HTCondor permet aux expériences du LHC (Atlas, Alice, CMS et LHCb) de disposer des ressources sur lesquelles l'IN2P3 s'est engagé. Ces évolutions lui permettent d'atteindre, sur les quatre dernières années, une disponibilité de 99% et une fiabilité de 99%.

La plateforme de calcul « locale » représente quant à elle 30% de la capacité totale proposée par le CC-IN2P3. Dernière plateforme de calcul utilisant le logiciel commercial d'ordonnancement Univa Grid Engine, le remplacement de ce dernier a nécessité un travail important de sélection et d'intégration initié en 2021. Le logiciel Slurm a été sélectionné et après une transition en douceur, malgré un calendrier très serré, il est en production depuis le premier semestre 2022.

La transformation des moyens de calcul du CC-IN2P3 en deux plateformes indépendantes s'est terminée avec un résultat très positif, au-delà même des attentes. Une économie de plus de 500 000 € sera réalisée au cours des cinq prochaines années avec l'abandon de la solution commerciale précédemment utilisée.

Afin de faciliter l'administration des deux plateformes de calcul, l'équipe a développé un système complet de surveillance et de comptabilité, offrant une vue de l'activité en temps réel, permettant également de consulter l'historique grâce à une interface web conviviale. Les données et les statistiques qu'elle fournit font partie intégrante de l'activité de gestion quotidienne des plateformes et de la prise de décisions à long terme.



Réseau

L'accroissement constant des besoins en terme bande passante réseau tant au sein du Centre de Calcul de l'IN2P3 que vers l'extérieur avec ses partenaires, continue de diriger les évolutions des réseaux locaux des salles serveurs et des réseaux étendus.

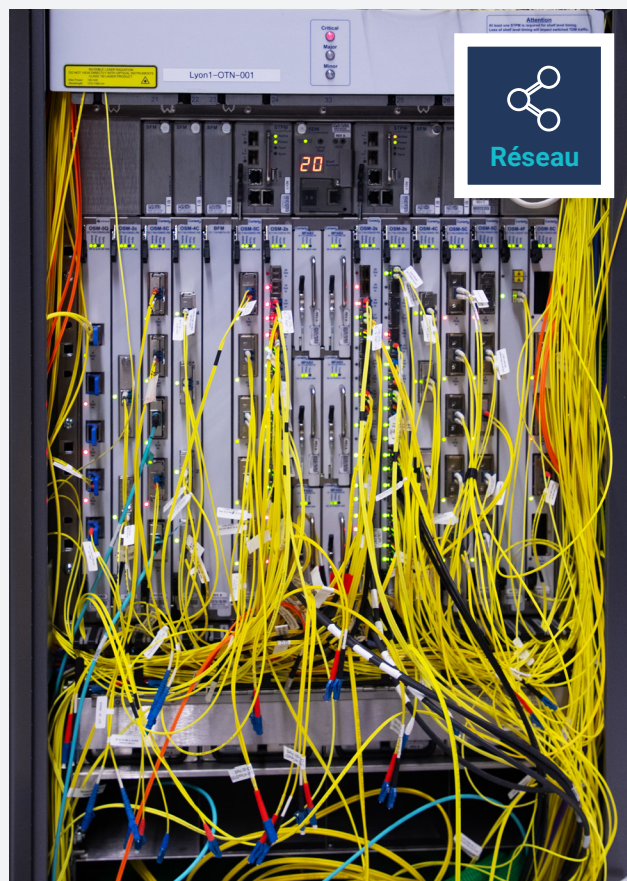
Pour fournir plus de ressources accessibles aux utilisateurs de la grille de calcul WLCG du LHC, mais également aux autres expériences, l'ensemble des serveurs de stockage connectés chacun à 10 Gb/s, est dorénavant agrégé sur des commutateurs remontant via de multiples liens 100 Gb/s vers l'épine dorsale du réseau local. Avec cette évolution, le nécessaire remplacement du routeur du cœur de réseau de la seconde salle serveurs, a été réalisé au cours de l'année 2023.

Ces dernières années, cette augmentation de bande passante a été également réalisée sur les nœuds de la plateforme de calcul, dont les connexions réseau ont progressivement évolué de 1Gb/s à 10Gb/s au gré des renouvellements des matériels. Des évolutions de l'infrastructure réseau sont en cours et d'autres à venir pour faire face à la demande croissante des expériences.

Pour les liaisons externes, le CC-IN2P3 dispose de multiples liens 100 Gb/s :

- A destination de l'IDRIS,
- Pour se raccorder au réseau LHCOPN des centres de calcul Tiers1 de WLCG (CERN), mais également au réseau LHCONE des sites WLCG et vers le réseau généraliste RENATER.

Pour répondre à la montée en puissance des expériences du LHC et face à l'utilisation mutualisée croissante du réseau LHCONE par d'autres expériences non-LHC (Belle II, DUNE, JUNO...), les liens 100 Gb/s pour le réseaux LHCOPN et LHCONE vont être doublés d'ici février 2024 et continueront à évoluer dans les prochaines années. A ce jour, ce sont plus de 400 To de données qui sont échangés chaque jour avec les sites partenaires du CC-IN2P3.



Localement, pour remplacer une infrastructure réseau vieillissante dans le bâtiment des bureaux, un projet de refonte complète a été lancé en 2022.

L'année 2023 a vu le renouvellement intégral des équipements du réseau sans fil, vieux de plus de 10 ans et le remplacement d'une desserte réseau reposant sur un câblage en cuivre obsolète. Une dorsale en fibre (FTTO), desservant les bureaux et les salles de réunions, a remplacé ce câblage vieux, présent pour certains tronçons depuis la création du bâtiment en 1986. Cette nouvelle infrastructure assure au personnel du CC-IN2P3, la pérennité d'un service réseau de qualité pour les dizaines d'années à venir.

Accompagnement

Le support apporté aux utilisateurs

L'équipe Accompagnement créée en novembre 2022 constitue l'interface entre les moyens du centre de calcul et les utilisateurs au sens large (expériences, projet, utilisateurs). Cette interface très majoritairement construite autour de relations humaines fonctionne dans deux directions : la première vise à assurer le bon usage des ressources par les utilisateurs et la seconde vise à répondre au mieux aux besoins des usagers.

Cet accompagnement dans l'usage et dans l'expression du besoin est au centre des missions de l'équipe ; missions qui se déclinent selon quatre axes :

L'exploitation opérationnelle consiste à gérer tous les aspects opérationnels liés aux services proposés aux utilisateurs. Il s'agit ni plus ni moins d'un service de support aux utilisateurs. Des personnels de l'équipe assurent non seulement une permanence opérationnelle pour surveiller la disponibilité des services informatiques mais aussi un suivi et une gestion des tickets d'incidents remontés par les utilisateurs.

L'accompagnement des expériences correspond principalement à une implication active dans la façon dont les utilisateurs interagissent avec les services du CC-IN2P3 notamment dans la définition de leurs modèles de calcul. C'est dans cette activité que les compétences des membres de l'équipe vont être principalement exploitées pour apporter aux utilisateurs le savoir-faire et le « savoir-utiliser » du centre de calcul. Des interactions fortes et resserrées sont alors nécessaires entre utilisateurs et membres de l'équipe.

L'intégration de nouveaux besoins est une activité permettant l'intégration de nouveaux services. La pertinence est avérée pour répondre à un besoin utilisateur. Cette activité participe à faire évoluer le portfolio des services qui sont mis à la disposition des utilisateurs du centre.

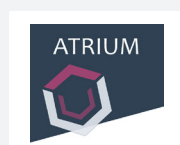
Le quatrième axe est beaucoup plus prospectif et consiste à faire **le suivi de projets de recherche et développement** dont les éventuels débouchés pourraient conduire le centre de calcul à mettre en œuvre de nouveaux services. Dans cette démarche l'équipe peut être amenée à participer, suivre, piloter, coordonner différents types de projets techniques qui s'inscrivent dans un cadre dont la dimension peut être institutionnel, national ou européenne.

L'élément moteur des actions menées au sein de l'équipe Accompagnement, reste les ponts et les relations humaines établies entre le centre de calcul et ses usagers.

Les outils

Les solutions logicielles pour répondre aux besoins des utilisateurs

Le centre de calcul propose des services collaboratifs à l'ensemble du personnel mais également aux collaborateurs externes à l'Institut. La dimension internationale des activités de l'IN2P3 a conduit le CC-IN2P3 à ouvrir ces outils à l'ensemble de la communauté Recherche (chercheurs, personnels techniques, étudiants) représentée dans la fédération d'identité eduGain.



Atrium

Plateforme de gestion électronique de documents basée sur le logiciel Nuxeo.



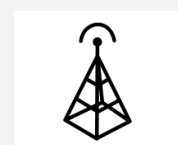
IN2P3-Forge

Plateforme de gestion de projet collaborative basée sur le logiciel Redmine.



Gitlab

Plateforme de gestion de projets en développement logiciel orienté DevSecOps.



Etherpad

Plateforme "bloc notes" collaborative.



Survey

Plateforme de gestion d'enquêtes et de sondages basée sur le logiciel LimeSurvey.



Messagerie électronique

Plateforme de messagerie électronique collaborative basée sur le logiciel Zimbra.



Box

Plateforme collaborative de gestion et d'édition de documents.



Chat

Plateforme de messagerie instantanée institutionnelle.

04.

Un savoir faire adapté aux expériences scientifiques



Physique des particules

LHC (Large Hadron Collider)

L'étude de l'infiniment petit passe notamment par l'analyse des collisions de particules que réalise le LHC au CERN. Le traitement des données issues des quatre expériences (ALICE, ATLAS, CMS et LHCb), qui compte cette installation, a nécessité la mise en œuvre de la grille de calcul WLCG (World LHC Computing Grid). Cette collaboration agrège les capacités de plus d'une centaine de sites à travers le monde.

Le projet LCG-France est la composante française de cette collaboration et coordonne la participation des sites français. L'IN2P3 s'est engagé à mettre à disposition des moyens de calcul et de stockage à hauteur de 10 % des ressources de WLCG et s'appuie principalement sur le CC-IN2P3, qui est un des centres de premier niveau à supporter les quatre expériences.

Les quatre dernières années ont été marquées par la pandémie du COVID qui a entraîné un décalage d'un an du calendrier de fonctionnement du LHC avec un arrêt de maintenance plus long. La troisième phase de prise de données, qui a démarré début 2022, est importante car elle prépare sur certains points opérationnels et technologiques les choix qui seront faits pour le passage du LHC au High Luminosity LHC, mise à jour majeur de l'accélérateur prévue à partir de 2026.

L'envol à venir du volume des données exige une nouvelle façon d'utiliser le stockage pour l'accès aux données avec une utilisation plus intense des bandes magnétiques. Le CC-IN2P3 a donc pris part à un challenge en octobre 2021 autour du stockage (disque et bande) et du réseau pour valider l'infrastructure de production mise en place.



Le Centre de Calcul a passé avec succès ce test en atteignant l'objectif des 10 % de bande passante requise pour le HL-LHC. Cela a notamment permis de mettre en lumière des insuffisances pour le monitoring sur lequel les équipes du Centre de Calcul ont travaillé. Cet exercice va être renouvelé fin février 2024 avec un objectif plus contraignant : 25 % de la bande passante maintenus pendant 48 h minimum, et participation simultanée des quatre expériences.

En parallèle, le CC-IN2P3 a participé à divers groupes de travail au sein de WLCG : transition des certificats vers les jetons pour l'authentification/autorisation, validation du modèle de Data Lake pour l'optimisation de la gestion et de l'accès aux données, élaboration d'une nouvelle suite pour l'évaluation de la puissance des serveurs de calcul.

Outre ces évolutions et l'adaptation de l'infrastructure, le CC-IN2P3 travaille continuellement pour maintenir la disponibilité et la fiabilité des services de grille au niveau des objectifs fixés par le memorandum d'entente que la France a signé avec WLCG. Il devra aussi être en mesure de fournir les ressources de calcul et de stockage fixés par cet engagement.

05.



Physique des particules

LSST (Large Synoptic Survey Telescope)

Dans le cadre de l'engagement envers le projet LSST pris par l'institut de contribuer au traitement des images de l'observatoire Vera C. Rubin, différentes actions ont été menées au cours de ces deux dernières années.

En premier lieu, une refonte complète de l'estimation des ressources informatiques nécessaires pour tenir ces engagements a été réalisée. Elle a permis de déterminer de façon plus précise le budget équipement nécessaire pour la période 2024-2035 grâce, d'une part à une meilleure compréhension du modèle de calcul envisagé pour traiter les images, et d'autre part, aux avancées récentes dans le logiciel développé par le projet Rubin spécifiquement pour cette fin. Ce budget prévisionnel de long terme a permis à l'IN2P3 d'adapter le niveau de sa contribution aux traitements à 40% des images brutes collectées pendant toute la durée de construction du relevé LSST.

Ces traitements se feront avec l'équipement localisé au CC-IN2P3 et avec une forte implication du personnel en étroite coopération avec les scientifiques de l'Institut participant au projet.

Des actions de consolidation de l'infrastructure informatique existante ont été conduites en tenant compte des besoins spécifiques du projet Rubin et des modifications ont été opérées en anticipation des besoins déjà identifiés. Plusieurs composants sont nécessaires au CC-IN2P3 pour la contribution à LSST et des travaux de préparation ont été faits et consolidés sur chacun d'entre eux.

Une infrastructure de stockage dédiée a été déployée, composée d'un système de stockage basé sur CephFS ainsi qu'une instance dédiée de dCache avec des capacités initiales autorisant des exercices de

taille conséquente et configurés pour croître sur long terme. Après des multiples exercices représentatifs, cette infrastructure a été dimensionnée en tenant compte des besoins spécifiques de LSST en termes de volumétrie, de modalités d'accès et du nombre de fichiers à accueillir. Le CC-IN2P3 a pris en charge le développement et la maintenance du composant logiciel utilisé par Rubin pour l'interaction avec dCache à travers le protocole webDAV ce qui nous permet d'avoir une connaissance détaillée sur toute la chaîne d'accès aux données, nous facilitant ainsi son optimisation.

La préparation de la ferme de calcul pour LSST a été également entamée. Des nœuds de calcul avec configuration spécifique (en termes de mémoire vive, disque local, carte réseau) ont été récemment intégrés en production dans la ferme Slurm et font l'objet d'une qualification. Un « computing element » basé sur ARC a été déployé afin d'exposer la capacité de traitement de la ferme Slurm à l'orchestrateur central de tâches de calcul de Rubin, basé sur PanDA. Une exploration vient d'être démarrée visant à identifier l'intérêt technique et économique de l'utilisation des nœuds de calcul sous Linux équipés de processeurs ARM.

Des bases de données PostgreSQL avec des dispositifs de stockage à faible latence ont été spécifiquement déployées aussi bien pour les besoins des campagnes annuelles de traitement des images que pour les besoins individuels des scientifiques membres de la communauté LSST France.

La configuration du réseau local a été modifiée afin d'éviter que le trafic induit par les traitements pour Rubin ne traverse les réseaux spécifiques du LHC (LHCONE et LHCOPN). Les serveurs de stockage et d'échange de données inter-site pour LSST ont été connectés sur des sous-réseaux séparés.

De ce fait, nous serons en capacité d'identifier plus finement les flux réseau à des fins de surveillance, par exemple.

Pendant une grande partie de l'année 2022 un exercice de traitement d'un jeu d'images simulées a été réalisé avec succès au CC-IN2P3. Dénommé le Data Preview 0.2, il visait à reproduire une campagne de traitement exécutée par le projet Rubin avec le même jeu de données en entrée en utilisant la plate-forme du cloud commercial de Google aux USA. Cet exercice a démontré la capacité du CC-IN2P3 à conduire et réussir une campagne d'envergure significative et nous a aidé à identifier les points d'amélioration nécessaires à apporter à l'infrastructure du site. Ce travail a été l'objet d'une présentation à la conférence CHEP 2023.

Nous avons également continué des explorations des outils d'analyse scientifique des données qui seront produites par Rubin. Une instance expérimentale de la Rubin Science Platform a été déployée au CC-IN2P3 : elle permet d'exposer des données résultantes de l'exercice Data Preview 0.2, en particulier le catalogue astronomique ingéré dans une instance locale de la base de données Qserv, aussi développée par Rubin. Bien que le budget pour l'équipement de la plate-forme d'analyse ne soit pas encore garanti par l'Institut, cette activité nous permet de nous préparer au cas où un budget serait dégagé pour cette activité essentielle pour la production scientifique des chercheurs de LSST France, d'une part, et elle nous permet également d'exposer dès maintenant les chercheurs français aux outils qui seront à leur disposition lorsque les données de Rubin seront publiées.

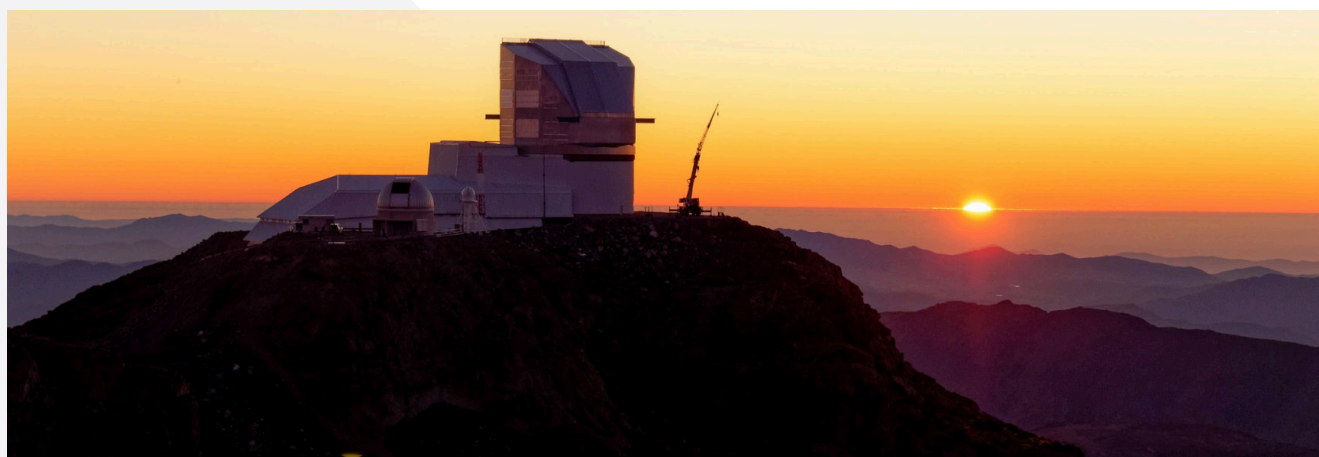
L'activité d'échange de données entre les sites participant au projet Rubin est centrale. Un groupe de travail a été créé au sein du projet et coordonné

par le CC-IN2P3 pour explorer et mettre en place des solutions de catalogage et de réplication automatisée de données. Ces solutions sont testées à travers une première campagne en cours qui vise à répliquer continuellement vers le CC-IN2P3 des images de tests du plan focal de la caméra actuellement localisée à SLAC. Cette activité est non seulement utile pour valider les solutions de réplication de données mais la présence de ces images au CC-IN2P3 permet aux chercheurs de l'Institut de participer confortablement à la qualification du plan focal auquel l'IN2P3 a fait une contribution majeure.

Le CC-IN2P3 continue d'être le site de référence pour la distribution des releases du logiciel de traitement des images de LSST via une infrastructure de distribution basée sur CernVM FS. Ce mécanisme de distribution du logiciel jusqu'aux sites de traitement est utilisé pour des exercices de traitement d'images avec des données d'autres instruments ou des données simulées conduits par le projet Rubin.

A l'initiative de LSST France, le CC-IN2P3 coordonne actuellement un projet qui vise à la caractérisation puis à la réduction de la consommation de ressources (principalement mémoire vive et CPU) du logiciel LSST. Avec le soutien financier de la Mission Interdisciplinaire du CNRS, nous sommes persuadés que ce travail pourra contribuer à réduire l'empreinte et le coût du projet Rubin dans son ensemble, avec un bénéfice direct pour l'IN2P3.

Plusieurs volets des travaux conduits pour LSST au CC-IN2P3 ont fait l'objet de présentations et de publications pour des conférences (CHEP2023, ADASSXXXIII) et ont constitué des fils conducteurs de la participation du CC-IN2P3 à des projets Européens tel ESCAPE.



© Rubin Obs./NSF/AURA/W. O'Mullane

Physique des particules

EUCLID

Euclid est une mission de l'ESA principalement dédiée à la cosmologie, qui a pour objectifs principaux de comprendre pourquoi l'expansion de l'Univers s'accélère sous l'effet d'une « énergie sombre », et de cartographier la « matière noire ». En observant toujours plus loin, donc en remontant toujours plus loin dans le temps, Euclid tentera de reconstruire l'évolution de notre Univers au cours des 10 derniers milliards d'années sous les effets antagonistes de ces deux composantes encore mystérieuses de notre Univers.

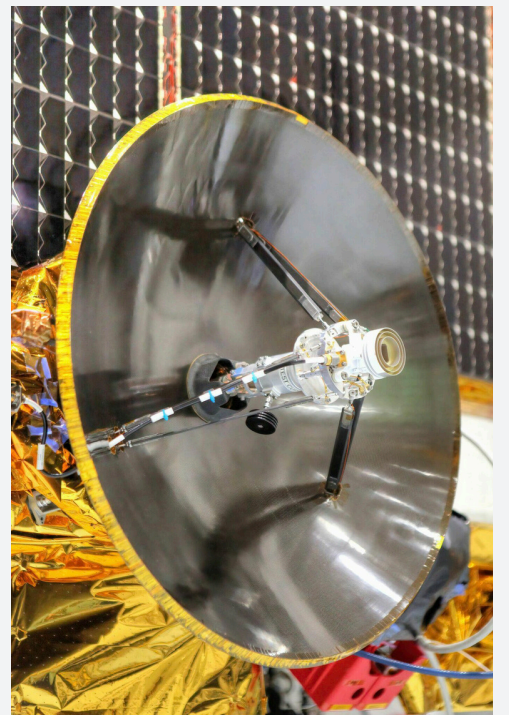
Le satellite Euclid a quitté la Terre le 01/07/2023 et doit observer à peu près un tiers de la voûte céleste au cours de sa mission nominale de 6 ans. Euclid observera donc des milliards de galaxies et l'évolution des grandes structures de l'Univers à travers les âges dans le domaine visible et proche infrarouge. Pour réaliser ce colossal travail de cartographie, Euclid emporte à son bord 2 instruments, un spectrophotomètre proche infrarouge appelé l'instrument NISP (Near Infrared Spectro Photometer) et un imageur travaillant dans le domaine visible, l'Instrument VIS (Visible Instrument), développés par un consortium international dirigé par la France. Le consortium Euclid regroupe plus de 2 200 personnes (dont 425 en France) réparties dans environ 250 laboratoires (dont 40 en France) de 16 pays.

Au cours de sa vie, la mission Euclid va fournir une quantité de données sans précédent pour une mission spatiale : environ 850 Gbit de données compressées par jour. Le segment sol scientifique du consortium Euclid est en charge du traitement des données reçues et de la fourniture des produits scientifiques. Les traitements sont réalisés dans 9 centres de traitement, 8 en Europe et 1 aux Etats-Unis. Pour la France, le centre de calcul de l'IN2P3 réalise à lui seul le stockage et le traitement de 30% des données Euclid. Les équipes du CC-IN2P3 sont responsables de la mise en place, de l'opération et de la maintenance des ressources de stockage, calcul et réseau nécessaires. Ces dernières années, le CC-IN2P3 a contribué aux développements de l'infrastructure du segment sol en participant activement aux activités préparatoires.

Des simulations de données ont été réalisées afin de permettre aux équipes du Consortium de mettre au point leurs programmes de traitement. Des campagnes de tests ont été exécutées dans le but de valider les pipelines logiciels ainsi que de vérifier la bonne montée en charge de l'infrastructure.

Les membres du Consortium ont également été accompagnés dans leur utilisation des services du Centre de Calcul. Des services dédiés à Euclid permettent une gestion automatisée des données et des traitements dans chaque centre impliqué. Ces services ont nécessité de nombreux tests en collaboration avec le consortium pour leur développement, et sont aujourd'hui opérés par les équipes du Centre de Calcul. La recette en vol du satellite s'est terminée le 5 octobre 2023 et la mission est ensuite entrée dans sa phase de calibration. Le CC-IN2P3 est aujourd'hui activement impliqué dans ces activités, en fournissant par exemple les ressources nécessaires aux traitements de calibration de l'instrument VIS. En fin de mission on estime qu'une capacité de calcul de plus de 70 000 HS06 sera nécessaires au CC-IN2P3 pour les traitements des données Euclid, équivalent à plus de 4000 coeurs de 2023. Plusieurs Po seront nécessaires pour stocker les données générées par ces traitements.

ESA/Euclid/Euclid Consortium/NASA, image processing by J.-C. Cuillandre
(CEA Paris-Saclay), G. Anselmi, CC BY-SA 3.0 IGO



Bilan de Gaz à Effet de Serre (BGES)

La démarche du Centre de Calcul

L'accord de Paris, adopté en 2015, prévoit de limiter les émissions de gaz à effet de serre (GES) de 2030 à 60% de celles de 1990. Le CNRS s'étant engagé à suivre cet accord, a mené un premier bilan en 2019 qui se traduit par la mise en place d'un plan de transition bas carbone.

Quatre axes d'action ont ainsi été identifiés :

- les achats
- le numérique
- la mobilité
- l'énergie

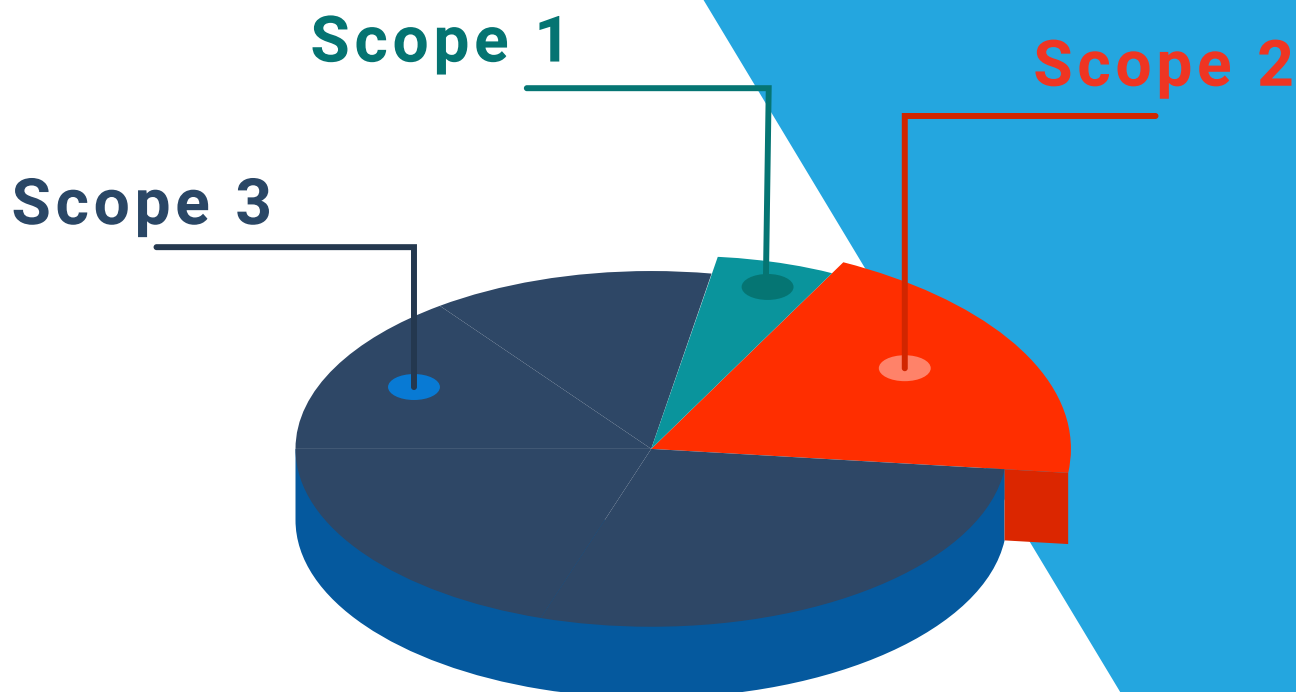
Dans ce contexte, il est essentiel de mesurer et d'améliorer en continu la performance environnementale des datacentres qui sont naturellement de gros consommateurs d'énergie. Le CC-IN2P3 est directement concerné par ce plan. Ce souci de l'efficacité du Centre de Calcul n'est pas récent. Etablie depuis 2010, la mission d'urbanisation des salles machines, transverse aux équipes informatiques et au service technique et logistique, a pour responsabilité de garantir une efficacité optimale des installations.

Plusieurs actions ont été menées ces dernières années, à commencer par la participation à un sous-groupe de travail mis en place par le Ministère de l'Enseignement Supérieur et de la Recherche, spécialisé pour les Infrastructures de Recherche Calcul. L'enjeu était de trouver les bons indicateurs de performance pour rendre un service de qualité dans une démarche durable, au-delà de suivre les normes classiques mises en place.

Un travail de caractérisation de facteurs d'émissions de gaz à effet de serre par service est également en cours. Il requiert la prise en compte des coûts carbone de la production et du transport de nos équipements, ainsi que de leur consommations électrique. Afin de maîtriser la facture énergétique, les leviers identifiés pertinents sont de se focaliser sur le matériel informatique, sur l'organisation du système d'information et sur l'efficacité des groupes froids et de la distribution électrique.

En terme de développement durable, la Recherche aura toujours besoin d'outils informatiques. Pour une recherche durable, le Centre de Calcul se doit donc d'être profondément impliqué et cherche la meilleure organisation possible.





57% des émissions de Gaz à effet de serre du CC-IN2P3 comprend non seulement ces achats informatiques, mais aussi l'énergie nécessaire comme l'électricité, pour entretenir les serveurs, ce qui en fait un cas à part.

Pour comptabiliser et gérer les émissions de GES, Le GHG Protocol (Green House Gas Protocol) est la norme internationale utilisée. Ce protocole définit trois catégories d'émissions, appelées "scopes" :

Le Scope 1 : les émissions directes de GES provenant des sources contrôlées par l'organisation, telles que les émissions de dioxyde de carbone (CO₂), de méthane (CH₄) et de protoxyde d'azote (N₂O) résultant des combustibles fossiles brûlés dans les installations de l'organisation, ou des émissions de CH₄ et N₂O provenant des procédés industriels.

Le Scope 2 : les émissions indirectes de GES provenant de la production d'électricité, de chaleur ou de vapeur achetées par l'organisation.

Le Scope 3 : toutes les autres émissions indirectes de GES qui résultent des activités de l'organisation, telles que les émissions de GES liées aux matières premières et aux produits achetés par l'organisation, les émissions de GES liées au transport des biens et des personnes, les émissions de GES résultant des déchets produits par l'organisation, etc.

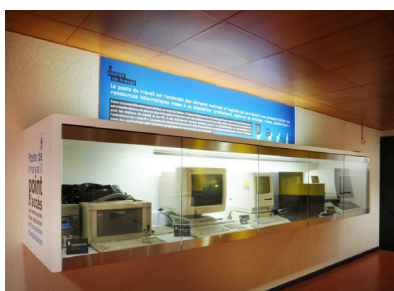
La majorité de la consommation du CC-IN2P3 regroupe le scope 1, le scope 2 et les achats IT (voir schéma).

06.

Patrimoine informatique

Vers un musée national de l'informatique ?

Engagé depuis 2011 dans la conservation et la valorisation de son patrimoine informatique, le CC-IN2P3 poursuit cette démarche et va plus loin avec un élargissement de ce projet à d'autres entités, au sein du CNRS et en dehors. Jusqu'à la création d'un musée national de l'informatique ?



Il existe en France un riche patrimoine scientifique, technologique et industriel à la fois matériel et immatériel qui témoigne d'une riche histoire de l'informatique. La conservation de ce patrimoine est aujourd'hui morcelée, assurée par des collectionneurs privés, des associations, des entreprises et des institutions. Ces collections constituent un ensemble exceptionnel de dimension internationale.

Les politiques d'acquisition de chacun se complètent, mais sont limitées dans leurs ambitions, ne disposant pas toujours des moyens financiers adaptés. La conservation de ces artefacts et leur valorisation auprès des publics sont en fonction du temps que chacun peut y consacrer mais pas à la hauteur de leur importance historique et culturelle. L'informatique, à la fois science, technologie, industrie et objet sociétal doit être prise en compte dans sa globalité. Aucune institution en France n'en a, à ce jour, assumé la mission.

Depuis 2011, le CC-IN2P3 s'est engagé dans une action de conservation de son patrimoine informatique avec la création d'un musée de l'informatique, une des rares initiatives de ce type en France. Visité par plusieurs centaines

de personnes par an, uniquement lors de visites guidées et privées, cet espace est devenu au fil du temps un élément central de la politique de médiation du Centre.

Celle-ci a pour objectif de rappeler le rôle grandissant de l'informatique dans les expériences de physique et d'expliquer les concepts de base du numérique à un public souvent plus intéressé par ses applications que par son fonctionnement. Elle a également vocation à montrer la diversité des métiers informatiques au CNRS dans un marché de l'emploi très concurrentiel.

Plus récemment, le CC-IN2P3 s'est engagé dans la création d'un consortium d'organismes académiques, d'associations et d'entreprises dont l'objectif est de faire émerger une initiative nationale de conservation du patrimoine allant jusqu'à l'idée de création d'un musée national de l'informatique et du numérique en France.

Le CC-IN2P3 participe aux discussions au sein de ce consortium aux côtés d'autres acteurs du CNRS (en particulier du LPNHE, de l'Observatoire de Paris et du Centre Alexandre Koyré).

07.

Patrimoine informatique

Les différents acteurs du projet disposent chacun d'une collection associée à des compétences spécifiques, allant de la muséographie, de la conservation de matériels à la médiation en passant par l'étude de l'histoire de l'informatique.

Le CC-IN2P3 contribue au projet avec sa propre collection de matériels, ses connaissances historiques, notamment sur le rôle que l'IN2P3 a joué dans l'émergence de certaines technologies, et son expérience en matière de médiation scientifique. Partie prenante dans les réunions de projet, il s'est lancé dans un inventaire exhaustif de ses collections et dans le développement d'une base de données ayant vocation à être intégrée à celle du projet Patstec (<https://www.patstec.fr>) où se trouvent déjà plusieurs acteurs du projet.

Le CC-IN2P3 a également entrepris de fédérer d'autres entités du CNRS autour de cette action de conservation du patrimoine informatique. Fort de son expérience dans ce domaine, le CC-IN2P3 est en effet convaincu qu'une telle initiative au sein du CNRS permettrait de positionner ce dernier comme acteur majeur du projet MINF aux côtés d'autres organismes

académiques d'importance (Aconit, ATOS, BnF, CNAM, CNRS, INRIA, Musée virtuel de Bull et de l'informatique Française, Hommo Calculus, Orange, MO5.COM, Société Informatique de France, Software Heritage, WDA, etc.).

Elle lui permettrait également de valoriser ses collections autant que les recherches en histoire de l'informatique qui sont faites en son sein (notamment au sein de INSHS), d'expliquer l'utilisation intensive qui en est faite par les utilisateurs du domaine (notamment par les chercheurs impliqués dans les grands programmes scientifiques de l'IN2P3) et de contextualiser et promouvoir les recherches en informatique réalisées à l'INS2I. Le CNRS trouverait ici une vitrine à l'échelle nationale et, confortant son expertise en matière de médiation, il contribuerait également à promouvoir les filières informatiques, notamment pour les publics jeunes et étudiants.

Les journées du patrimoine 2023 //

Signe que le sujet passionne, le samedi 16 septembre 2023, près de 900 personnes ont visité le CC-IN2P3 lors des Journées Européennes du Patrimoine.

Au programme :

Visites des salles informatiques et du musée, exposition photographique « Du calcul aux données », focus sur la naissance du Web avec l'intervention de Daniel Chamay, pionnier du web, mise à disposition d'un espace jeux sur le Rétrogaming proposé en collaboration avec l'association G2L2 Corp., conférence sur l'histoire de l'informatique par Pierre-Eric Mounier-Kuhn, historien du CNRS, et Damien Bourelle, membre d'Aconit.



© CC-IN2P3 / CNRS

Le premier serveur Web français : <http://info.in2p3.fr> a été mis en place sur la Station NeXTcube



Adresse

Centre de Calcul de l'IN2P3
21, avenue Pierre de Coubertin
Campus Lyon-Tech-la Doua
69100 Villeurbanne



Téléphone

Tél. : +33 (0)4 78 93 08 80



Site Web

www.cc.in2p3.fr



Contact

contact@cc.in2p3.fr