



CSCS

Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

Lugano, 11 juillet 2018

Stockage de données en pétaoctets au CSCS pour l'Institut Paul Scherrer

Les données de recherche recueillies par les grandes installations de recherche de l'Institut Paul Scherrer (PSI) à Villigen seront à l'avenir archivées au Centre suisse de calcul scientifique (CSCS) à Lugano. Ceci est rendu possible grâce à une collaboration entre le PSI et le CSCS.

L'enregistrement et l'archivage des données font partie des activités quotidiennes du CSCS. Mais le transfert en ligne par câbles à fibres optiques à travers toute la Suisse d'énormes quantités de données relatives à de grands projets de recherche en cours constitue un véritable défi logistique. Il en est de même pour le transfert d'archives de données complètes couvrant plusieurs décennies. C'est ce que révèle un projet de l'Institut Paul Scherrer (PSI) à Villigen et du CSCS à Lugano. Le PSI et ses chercheurs utilisent depuis près de 15 ans les ordinateurs ultraperformants du CSCS qui fait partie de l'EPF Zurich. Désormais, le PSI enregistre également ses données au CSCS – des données générées par les grandes installations de recherche du nouveau laser à rayons X SwissFEL et la Swiss Light Source (SLS). Au total, il s'agit de 10 à 20 pétaoctets de données par an. A titre d'illustration: pour stocker ne serait-ce qu'un pétaoctet de données sur DVD, il faudrait en utiliser plus de 200 000.

Un transfert de données en toute sécurité

Le transfert des données générées au PSI résultant directement de l'expérimentation est désormais possible grâce au projet récemment achevé avec succès par les collaborateurs du CSCS et du département Informatique du PSI. Le PSI et le CSCS ont notamment développé conjointement un processus permettant de transférer, d'archiver, de récupérer ou même, après expiration du délai d'archivage, de supprimer les données en toute sécurité – une sorte de guide sur la manière dont les données, à l'instar de colis, doivent être emballées pour être expédiées en toute sécurité. Avec les «spécifications d'emballage» développées à cet effet, ils ont intégré dans le processus de transmission, de stockage et de migration des données un workflow manager intelligent qui coordonne le processus. Ils ont en outre introduit un catalogue de données dans le processus d'archivage, qui liste systématiquement les données et facilite ainsi l'accès à ces dernières. Une connexion réseau redondante propre dédiée entre le PSI et le CSCS permet désormais de transférer 10 gigabits de données par seconde. La nouvelle solution est un atout majeur pour le CSCS et le PSI pour répondre aux demandes croissantes dans le domaine de l'ouverture des données de recherche.

Les initiateurs du projet se félicitent de la réussite du projet. «Ce projet est pour nous une évolution logique des services que nous proposons à nos utilisateurs», déclare Thomas Schulthess, professeur à l'EPF et directeur du CSCS. «Au CSCS, notre infrastructure informatique nous permet en outre de disposer des ressources nécessaires pour analyser ces données en fonction des besoins. Maintenir le stockage, l'analyse et le traitement des données dans un même lieu permet d'économiser du temps et de l'énergie.» Au PSI, on se réjouit pareillement de cette collaboration fructueuse. «A l'ère du big data et des projets de recherche avec des volumes de données importants, nous considérons comme une conséquence logique le fait de transférer nos données au CSCS et d'économiser ainsi des coûts par rapport à une solution locale», explique Gerd Mann, directeur du département Informatique au PSI. «La



CSCS

Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

coopération avec le CSCS permet de concentrer les ressources informatiques du PSI sur l'évaluation scientifique des données et sur de nouveaux projets.»

Le CSCS dispose d'une bibliothèque de bandes avec une capacité de stockage actuelle de 120 pétaoctets pour l'archivage des données des utilisateurs. D'ici 2022, le PSI transférera près de 85 pétaoctets de données au CSCS afin que ces données y soient archivées. Près de 35 pétaoctets proviennent des expérimentations du SwissFEL et 40 pétaoctets de la SLS. Au cours des cinq prochaines années, la capacité de stockage pourra être portée à 2000 pétaoctets afin de répondre au besoin croissant de stockage et d'archivage de données en Suisse.

Photos:

<https://www.cscs.ch/publications/press-releases/2018/589/>

Contact CSCS

Dr. Valentina Tamburello, Communications Officer

Phone: +41 78 964 36 40

Email: communication@cscs.ch

Dr. Michele De Lorenzi, Deputy Director

Phone: +41 91 610 82 08

Email: michele.delorenzi@cscs.ch

Contact PSI

Dr. Gerd Mann, Head IT PSI and Head IT Department

Phone: +41 79 799 62 98

Email: gerd.mann@psi.ch