



CSCS

Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

Lugano, 11. Juli 2018

Petabyte-Datenspeicher für das Paul Scherrer Institut am CSCS

Die Forschungsdaten, die die Grossforschungsanlagen des Paul Scherrer Instituts (PSI) in Villigen sammeln, werden künftig am Nationalen Hochleistungsrechenzentrum (CSCS) in Lugano gespeichert und archiviert. Ermöglicht hat dies eine Kollaboration zwischen PSI und CSCS.

Daten speichern und archivieren gehört zum Alltagsgeschäft des CSCS. Werden aber die enormen Datenmengen laufender Grossforschungsprojekte online per Glasfaserkabel quer durch die Schweiz transferiert, sowie ein ganzes, Jahrzehnte umfassendes, Datenarchiv, bedeutet das eine logistische Herausforderung. Dies zeigt ein Projekt des Paul Scherrer Instituts in Villigen und des CSCS in Lugano. Das PSI ist mit seinen Forschern seit rund 15 Jahren Benutzer der Hochleistungsrechner des zur ETH Zürich gehörenden CSCS. Neu speichert das PSI nun auch seine Daten am CSCS: Daten die die Grossforschungsanlagen des neuen Röntgenlaser SwissFEL und die Swiss Lightsource (SLS) generieren. Insgesamt handelt es sich pro Jahr um 10 bis 20 Petabyte an Daten. Zur Veranschaulichung: Wollte man nur ein Petabyte Daten auf DVDs speichern, bräuchte es davon mehr als 200'000 Stück.

Sicherer Datentransfer

Den Datentransfer der am PSI generierten Daten, die direkt vom Experiment kommen, ermöglicht nun das kürzlich erfolgreich abgeschlossene Projekt von Mitarbeitern des CSCS und der Abteilung IT des PSI. Zusammen entwickelten PSI und CSCS unter anderem einen Prozess, mit dem sich die Daten sicher transferieren, archivieren, wieder auffinden oder auch nach Ablauf der Archivierungsfrist löschen lassen – quasi eine Anleitung, wie Daten, analog zu Päckchen, für den sicheren Versand verpackt sein müssen. Bei den hierfür entwickelten «Verpackungsvorgaben» integrierten sie in den Prozess der Datenübertragung, -speicherung und -migration einen intelligenten Workflow Manager, der den Vorgang koordiniert. Darüber hinaus integrierten sie einen Datenkatalog in den Archivierungsprozess, der die Daten systematisch auflistet und somit den Zugriff darauf erleichtert. Über eine eigens eingerichtete dedizierte redundante Netzwerkverbindung zwischen PSI und CSCS lassen sich nun pro Sekunde 10 Gigabit Daten transferieren. Die neue Lösung ist für das CSCS und PSI ein wesentlicher Beitrag um die wachsenden Anforderungen im Bereich Open Research Data zu erfüllen.

Die Projektinitianten freuen sich über den erfolgreichen Abschluss. «Dieses Projekt ist für uns eine logische Weiterentwicklung der Services die wir unseren Benutzer bieten», sagt ETH-Professor Thomas Schulthess, Direktor des CSCS. «Zudem verfügen wir am CSCS mit unserer Rechnerinfrastruktur über die benötigten Ressourcen um diese Daten je nach Bedarf zu analysieren. Datenspeicherung, -analyse und -prozessierung nahe beieinander zu halten spart Zeit und Energie.» Auch seitens des PSI freut man sich über die geglückte Kooperation. «Für uns ist es eine logische Konsequenz im Zeitalter von Big Data aufgrund datenintensiveren Forschungsprojekte unsere Daten ans CSCS zu transferieren und gegenüber einer lokalen Lösung Kosten einzusparen», sagt Dr. Gerd Mann, Leiter Gesamt-IT am PSI. «Durch die Kooperation mit dem CSCS können die IT Ressourcen des PSI auf die Unterstützung der wissenschaftlichen Daten-Auswertung und neue Projekte fokussieren.»



CSCS

Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

Am CSCS steht für die Daten-Archivierung der Benutzer eine Tape Library mit einer aktuellen Speicherkapazität von 120 Petabyte zur Verfügung. Das PSI wird bis 2022 rund 85 Petabyte Daten ans CSCS transferieren und dort archivieren. Etwa 35 Petabyte stammen dabei von den Experimenten des SwissFEL und 40 Petabyte von der SLS. Im Laufe der nächsten fünf Jahren kann die Speicherkapazität bei Bedarf auf 2'000 Petabyte erhöht werden, um den wachsenden Bedarf an Datenspeicherung und -archivierung in der Schweiz zu sichern.

Fotos:

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

Kontakt CSCS

Dr. Valentina Tamburello, Communications Officer

Telefon: +41 78 964 36 40

Email: communication@cscs.ch

Dr. Michele De Lorenzi, Deputy Director

Telefon: +41 91 610 82 08

Email: michele.delorenzi@cscs.ch

Contact PSI

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

Telefon: +41 79 799 62 98

Email: gerd.mann@psi.ch