idw - Informationsdienst Wissenschaft
Nachrichten, Termine, Experten

# Pressemitteilung

Max-Delbrück-Centrum für Molekulare Medizin in der Helmholtz-Gemeinschaft
Annette Tuffs

14.09.2017
http://idw-online.de/de/news681016

Forschungsprojekte
fachunabhängig
überregional

# Wissenschaftliches Rechnen erfolgreich reproduzieren

**Um während eines Experiments Datenberge erheben und später interpretieren zu können, ist die Wissenschaft zunehmend von computergestützten Analyseverfahren und Hochleistungsrechnern abhängig. Unterschiede in den Rechenumgebungen verursachen jedoch Probleme, wenn andere Gruppen die Ergebnisse reproduzieren wollen. Ein Team vom Berliner MDC arbeitet an einer Lösung.**

Ein Eckpfeiler der Wissenschaft ist, dass Experimente und Ergebnisse reproduzierbar sein müssen. Soweit die Theorie. In der Praxis ist dieses Ziel mitunter schwer zu erreichen. Denn gerade die experimentellen Rahmenbedingungen moderner Hochdurchsatzverfahren sind nicht nur teuer, sondern das Resultat jahrelanger äußerst sorgfältiger Arbeit. Dass komplexe, maßgeschneiderte Computerprogramme oft die Analyse und Interpretation von Daten übernehmen, macht es noch schwerer, die Ergebnisse einer Forschungsgruppe unabhängig nachzuvollziehen.

Guix – eine freie Software, die Rechenumgebungen exakt und vollständig reproduzieren kann – könnte bei der Lösung dieses Problem helfen, sagt Ludovic Courtès vom Nationalen Forschungsinstitut für Informatik und Automatisierung in Frankreich (Inria). Er arbeitet zusammen mit Ricardo Wurmus vom Berlin Institute for Medical Systems Biology (BIMSB) am Max-Delbrück-Centrum für Molekulare Medizin, sowie Kolleginnen und Kollegen des Utrecht Bioinformatics Center am Universitätsklinikums Utrecht und anderen Institutionen daran, Guix in den Dienst der wissenschaftlichen Reproduzierbarkeit zu stellen.

Erfassung kompletter Software-Umgebungen

Die National Science Foundation in den USA und Fachzeitschriften wie Nature legen auf die freie Verfügbarkeit von Quellcode und größtmögliche Reproduzierbarkeit mittlerweile großen Wert. „Doch ob ein Experiment reproduzierbar ist, hängt auch davon ab, ob sich die Softwareumgebung reproduzieren lässt“, sagt Courtès. „Besonders schwierig wird es, wenn Hochleistungsrechner erforderlich sind.“

Guix, das Teil des GNU-Projektes ist, löst mehrere Probleme auf einen Schlag. Es macht Nutzer von der Software-Verwaltung durch Administratoren unabhängig, sie können Softwareumgebung ganz an ihre Bedürfnisse anpassen. Das Programm hat auch Vorteile gegenüber beliebten „Containerlösungen“, die vergleichbar sind mit einem neuen Computer, auf dem bereits alles vorinstalliert ist, sagt Courtés. „Das funktioniert nur so lange, bis kleine experimentelle Änderungen vorgenommen werden, um eine neue Hypothese zu testen. Und das geschieht häufig in der Forschung.“

Die Guix-Software basiert auf der Idee, Software-Umgebungen eindeutig zu beschreiben, ähnlich einer mathematischen Funktion. So lassen sich alle Beziehungen von Programmen untereinander charakterisieren und diese auf das Bit genau reproduzieren. Auf diese Weise vereinigt Guix zwei Anforderungen an eine wissenschaftliche Software-Umgebung: Reproduzierbarkeit und individuelle Anpassung.

Guix für Wissenschaftler

Allerdings wurde Guix ursprünglich nicht für das wissenschaftliche Hochleistungsrechnen (engl. high performance computing, HPC) konzipiert. Daher arbeiten Wissenschaftlerinnen und Wissenschaftler von MDC, Inria und ihren Partnern an Funktionen, die Guix auf Rechen-Clustern lauffähig machen, um auch dort reproduzierbare Arbeitsabläufe zu etablieren. Außerdem passen sie Programme an das Guix-System an und stellen sie als Pakete zur Verfügung.

„Vor Guix war die Installation von wissenschaftlicher Software zwangsläufig ad-hoc", sagt Ricardo Wurmus. „Forschungsgruppen haben ihre eigenen Softwarevarianten kompiliert, sie statisch in bestehenden Systemen verlinkt und gehofft, dass sich nichts ändert. Denn die Verwaltung von Softwareumgebungen war praktisch unmöglich. Heute managen wir mit Guix nicht nur eine Umgebung je Forschungsgruppe. Wir verwenden es auf allen Ebenen: Gruppe, Benutzer, Workflow und so weiter."

Das Projekt soll zwei Jahre laufen. In dieser Zeit wollen sie Software-Reproduzierbarkeit an ihren Institutionen erreicht haben. Courtès sagt: „Mit Guix machen wir einen großen Schritt hin zu einem grundlegen wissenschaftlichen Ziel. Davon wollen wir alle überzeugen, die auf Hochleistungs-Rechner angewiesen sind."

Das Max-Delbrück-Centrum für Molekulare Medizin (MDC)

Das Max-Delbrück-Centrum für Molekulare Medizin in der Helmholtz-Gemeinschaft (MDC) wurde 1992 in Berlin gegründet. Es ist nach dem deutsch-amerikanischen Physiker Max Delbrück benannt, dem 1969 der Nobelpreis für Physiologie und Medizin verliehen wurde. Aufgabe des MDC ist die Erforschung molekularer Mechanismen, um die Ursachen von Krankheiten zu verstehen und sie besser zu diagnostizieren, verhüten und wirksam bekämpfen zu können. Dabei kooperiert das MDC mit der Charité – Universitätsmedizin Berlin und dem Berlin Institute of Health (BIH) sowie mit nationalen Partnern, z.B. dem Deutschen Zentrum für Herz-Kreislauf-Forschung (DHZK), und zahlreichen internationalen Forschungseinrichtungen. Am MDC arbeiten mehr als 1.600 Beschäftigte und Gäste aus nahezu 60 Ländern; davon sind fast 1.300 in der Wissenschaft tätig. Es wird zu 90 Prozent vom Bundesministerium für Bildung und Forschung und zu 10 Prozent vom Land Berlin finanziert und ist Mitglied in der Helmholtz-Gemeinschaft deutscher Forschungszentren. http://www.mdc-berlin.de/

URL zur Pressemitteilung: https://guix-hpc.bordeaux.inria.fr/ Die Website des Guix-HPC-Projekts

URL zur Pressemitteilung:
https://insights.mdc-berlin.de/de/2017/09/wissenschaftliches-rechnen-erfolgreich-reproduzieren/ Diese Meldung bei MDC Insights

URL zur Pressemitteilung:
https://www.inria.fr/en/centre/bordeaux/news/towards-reproducible-software-environments-in-hpc-with-guix Detaillierte Informationen in der Pressemitteilung des Nationalen Forschungsinstitut für Informatik und Automatisierung in Frankreich (Inria)

Data Centre am MDC.
Bild: MDC