



ALOMA: Ein Parallelisierungsframework – nicht nur für seismische Anwendungen

Komplexe Rechenoperationen auf wachsenden Datenmengen korrekt auszuführen, ist eine der Herausforderungen von Big Data. Gemeistert werden kann sie nur mit der massiven Parallelisierung von Rechnern beziehungsweise Rechenleistung.

Die Forschenden der Abteilung »High Performance Computing« nutzen bereits seit Jahren GPI-Space als Programmier-Plattform für Hochleistungssysteme. Das Besondere ist die Userfreundlichkeit: Kundinnen und Kunden müssen sich kein spezielles HPC-Wissen aneignen, denn das Softwaresystem übernimmt die effiziente Ausführung der Algorithmen. Das gilt auch für die spezialisierte Version ALOMA.

ALOMA erkennt Datenabhängigkeiten

Eingesetzt wird das Tool meist in der Seismik. Konzipiert ist es aber als generelles Framework für die Ausführung von Workflows auf verteilten Systemen, denn ALOMA erkennt Abhängigkeiten in den Datensätzen, die als Input geliefert werden – unabhängig von der Datenquelle und beantwortet folgende Fragen: Wie werden die Daten verteilt? Welche können gleichzeitig und unabhängig voneinander bearbeitet werden? Wo können sie bearbeitet werden?

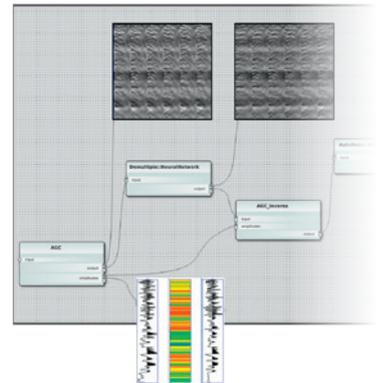
»Um Abhängigkeiten zu erkennen und diese Fragen zu beantworten, muss ALOMA entsprechende Informationen über die einzelnen Module und deren Kombination im Workflow zur Verfügung haben oder generieren. Für diesen Workflow verwenden wir eine Darstellung als Petri-Netz, welche die interne Workflow-Engine entsprechend analysieren kann«, erläutert Projektleiter Dr. Dirk Merten. Die Nutzenden müssen dazu nur in einem graphischen Workflow-Editor Ausgänge von algorithmischen Modulen

mit den Eingängen weiterer Module verbinden. Informationen über die einzelnen Module werden bei deren Einbindung in ALOMA angegeben. Diese Informationen beinhalten die Anzahl und Typen der Ein- und Ausgabedaten des Moduls sowie deren Granularität. Dies sind für die Seismik spezifische Werte (z. B. »Seismische Daten« oder »Geschwindigkeitsmodell«) sowie die Granularitäten »Trace«, »Gather« oder »Inline«.

Klassische Algorithmen und ML

Neben den für die Seismik typischen Granularitäten unterstützt ALOMA auch Datenaufteilungen aus der Mustererkennung und dem Maschinellen Lernen. Damit können ebenso Module eingebunden werden, die aus dem Deep Learning hervorgegangen sind. In einem exemplarischen Workflow kann ein Volumen von Eingangsdaten mehrfach korrigiert, zu einem Volumen gestapelt und auf Fehlermuster analysiert werden. Verarbeitungs- und Analysemodule für klassische Algorithmen und neuronale Netzinferenz werden einfach in einem Arbeitsablauf kombiniert. Die Parallelisierung aller Module innerhalb des Workflows übernimmt ALOMA automatisch.

Dank ALOMA können Seismik-Expertinnen und -Experten neue klassische oder auf Machine Learning basierende Algorithmen und Prototypen innerhalb kürzester Zeit einbinden, mit vorhandenen Algorithmen in Workflows kombinieren und parallelisiert auf realistischen Datenmengen und unter Alltagsbedingungen testen.



Ausschnitt eines exemplarischen Workflows mit ALOMA

Kontakt

Dr. Dirk Merten
Abteilung »High Performance Computing«
Telefon +49 631 31600-4616
dirk.merten@itwm.fraunhofer.de



Weiterführende Informationen gibt es auf unserer Website unter www.itwm.fraunhofer.de/aloma