



Sentinel-2A-Aufnahme von landwirtschaftlichen Flächen um den Limburger Hof bei Ludwigshafen am Rhein (22. 9. 2016)

Schnelle und effiziente Verarbeitung von Erdbbeobachtungsdaten

Mit dem Start des europäischen Copernicus-Programms wurden die Möglichkeiten im Hinblick auf Erfassung und Verfügbarkeit von Erdbbeobachtungsdaten erheblich erweitert. Durch Copernicus werden mehr Erdbbeobachtungsdaten als je zuvor aufgenommen, Aktualisierungszyklen werden stark verkürzt und alle Daten sind öffentlich frei verfügbar. Wie sieht es jedoch mit Zugriffsmöglichkeiten und Nutzbarkeit der Daten aus? Nur über einen nutzeroptimierten Zugang lässt sich auch die Wertschöpfung steigern. Hier steht die Branche vor einem IT-Paradigmenwechsel. Statt Desktop-basierten Fachanwendungen werden leistungsfähige und skalierbare Infrastrukturen für die Verarbeitung, Analyse und Visualisierung sehr großer Datenbestände aufgebaut. Die Algorithmen für die Analyse müssen effizient und skalierbar umgesetzt und es muss ein leichter Zugang zu Daten und Rechenkapazitäten gewährleistet werden.

Autoren: Dr. Michel Krämer und Philipp Willkomm

In den letzten Jahren hat sich in Hinblick auf neue Plattformen und Zugänge zu Prozessierungs- und Speicherkapazität jede Menge getan. Worauf kommt es bei der Bereitstellung von cloudbasierten Verarbeitungs- und Datenzugriffsdiensten an? Worauf sollten sowohl die Betreiber als auch die Nutzer dieser Angebote achten, insbesondere bei der Kombination der Copernicus-Daten mit eigenen Daten oder solchen aus anderen Quellen? In einem vom DLR geförderten Verbundprojekt haben sich das Fraunhofer IGD und die M.O.S.S. Computer Grafik Systeme GmbH mit genau diesen Fragestellungen beschäftigt.

Das Setting

Im Rahmen des Projekts wurde – aufbauend auf der Software „novaFactory“ – eine Ausführungsumgebung für eine effiziente Verarbeitung komplexer Datenauswertungen bereitgestellt. Neben dem Zugriff und der Verarbeitung von Satellitendaten aus dem Copernicus-Programm wurden im Rahmen einer Pilotanwendung noch weitere Daten und Verarbeitungsdienste eingesetzt. Die „novaFactory“-Ausführungsumgebung wurde in verschiedenen Cloudumgebungen eingebunden, die über ein definiertes Benchmarking miteinander verglichen und evaluiert wurden.

Die Pilotanwendung

In Abstimmung mit dem Hessischen Landesamt für Bodenmanagement und Geoinformation (HLBG) wurde eine praxisnahe Anwendung im Bereich Flächen-

monitoring definiert und umgesetzt. Im ersten Schritt dieser Pilotanwendung wurden aus bildbasierten digitalen Oberflächenmodellen (bDOM) sowie Luftbildern True-DOP berechnet und in einem zweiten Schritt folgte dann über mehrere Prozessschritte hinweg eine Änderungsdetektion, aus der sich Fortführungshinweise für die aktuelle Landnutzungsklassifizierung ableiten ließen. Eingangsdaten waren – neben den über die Code-DE-Plattform (Copernicus Data and Exploitation Plattform – Deutschland) des DLR verfügbaren Sentinel-2-Daten aus dem Copernicus-Programm – weitere Datensätze, wie zum Beispiel bildbasierte Oberflächenmodelle, Luftbilder und das aktuelle ATKIS-DLM als Referenzdatensatz.

Anforderungen an eine cloudbasierte Verarbeitungsplattform

Der Anwendungsfall setzte sich aus mehreren Prozessschritten zusammen, die nacheinander und teilweise parallel ausgeführt werden konnten, zwischen denen aber strenge Abhängigkeiten bestanden. Solche datengetriebenen Workflows (sogenannte „Scientific Workflows“) sind bei der Verarbeitung von Erdbeobachtungsdaten, aber auch anderer Geodaten sehr üblich. Möchte man in dieser Anwendung die einzelnen Prozessschritte nicht manuell verwalten, ist die Verwendung eines Workflow-Management-Systems ratsam. Spezialisierte Workflow-Management-Systeme sind in der Lage, die Ausführung der Arbeitsabläufe vollständig automatisiert zu überneh-

men und die Übertragung der Daten zu verwalten. Die im Projekt ausgewählte und eingesetzte Lösung „novaFactory“ steht beispielhaft für solche verteilten Ausführungsumgebungen für Scientific Workflows, ebenso wie das vom Fraunhofer IGD entwickelte Workflow-Management-System Steep.

Die oben beschriebene Pilotanwendung stellte besondere Herausforderungen an das System – insbesondere, was die Größe der zu verarbeitenden Daten anging. Die für den Anwendungsfall notwendigen Luftbilddaten deckten ca. die Hälfte des Bundeslands Hessen ab und hatten eine Größe von mehreren Terabyte. Eine Voraussetzung war deshalb, dass die Verarbeitungsplattform eine schnelle Netzwerk-anbindung, insbesondere eine möglichst direkte Anbindung an die Bilddaten des Copernicus-Programms besitzt. Eine Übertragung der Massendaten in die Cloudumgebung war durch Unterstützung des DLR über das Projekt Code-DE möglich.

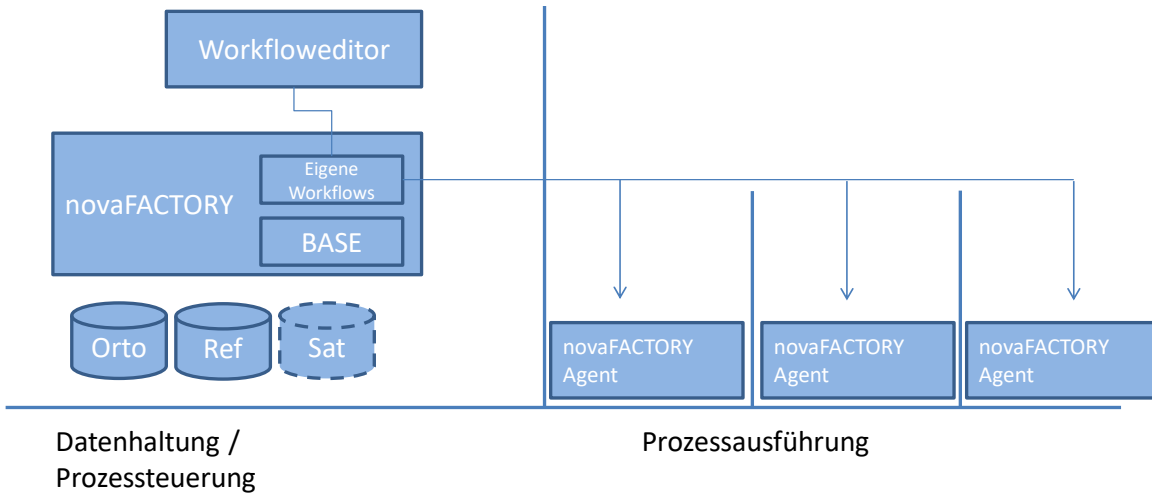
Außerdem sollte gewährleistet werden, dass in der Praxis bereits genutzte Dienste in die Verarbeitung einbezogen werden können. Ein dem konkreten Anwendungsfall zugrunde liegender Prozessbaustein setzte ein Windows-Betriebssystem voraus, sodass keine rein Linux-basierte Verarbeitungsplattformen genutzt werden konnte.

Neben technischen Vorgaben galt es insbesondere auch organisatorische Anforderungen an die Ausführungsumgebung (Datenhoheit etc.) zu berücksichtigen. Vor dem Hintergrund dieser Anforderungen



Die Arbeitsschritte der Pilotanwendung

Bild: M.O.S.S., Computer Grafik-Systeme GmbH



Übersicht über die Komponenten des entwickelten Verarbeitungssystems

wurde die Open Telekom Cloud (OTC) als Plattform favorisiert und im Projekt benutzt.

Evaluation

Mit der Kombination OTC und Code-DE konnte ein zu dem Zeitpunkt der Studie optimales Setting für die Ausführung des Anwendungsfalls mit „novaFactory“ identifiziert werden. Neben dieser Kombination wurden weitere Cloudumgebungen getestet sowie ein Vergleich mit den Leistungsparametern in einem „Reallabor“ vorgenommen. Das heißt, das Ausführungsverhalten wurde in einer Umgebung untersucht, wie

sie heute in der Regel bei Nutzern in Behörden und Unternehmen vorkommt.

Die Ergebnisse der Evaluierung haben gezeigt, dass Anwendungen aus dem Erdbeobachtungsbereich von Parallelisierung und Verteilung auf verschiedene Recheneinheiten profitieren und dass trotz Overhead und nicht zu 100 % linearer Skalierung ein großer Geschwindigkeitsvorteil erreicht werden kann.

In der Erprobung und der Einführung neuer und intelligenter Verfahren für die Verarbeitung von Erdbeobachtungsdaten und abgeleiteten Daten liegt aus Sicht der Verbundpartner ein klares wissenschaft-

liches und wirtschaftliches Potenzial. Die Erschließung dieser Verwertungspotenziale erfordert jedoch auch Orientierungshilfen und vergleichbare Kriterien für Anwender. Mit der vorgelegten Studie konnte ein Beitrag dazu geleistet werden.

Kontakt:

Dr. Michel Krämer
 Fraunhofer-Institut für Graphische Datenverarbeitung IGD
 E: michel.kraemer@igd.fraunhofer.de
Philipp Willkomm
 M.O.S.S. Computer Grafik Systeme GmbH
 E: pwillkomm@moss.de