

Big Data für die Erdbeobachtung

Ein grenzenloses Speichersystem für das EODC

Ziel:

Ein Speichersystem ohne Grenzen für das EODC zur Erdbeobachtung

Umsetzung:

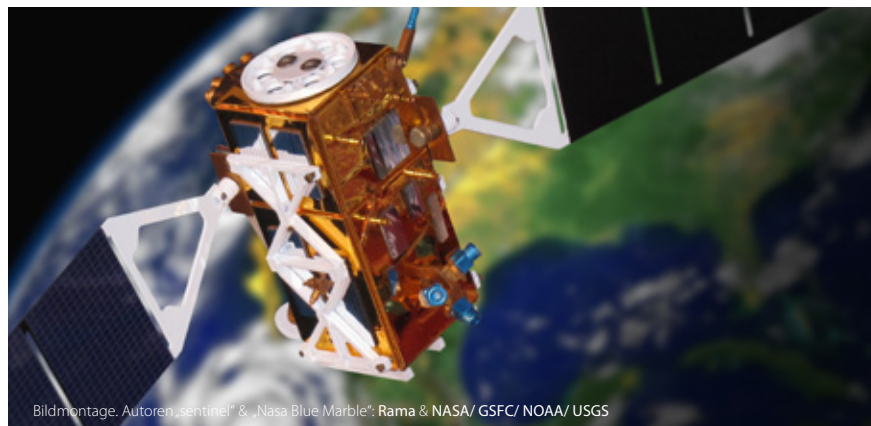
EDV-Design fungiert als Generalunternehmen: Planung, Präsentation, Aufbau und Adaption der Infrastruktur, Installation und Anbindung, Konfiguration und Inbetriebnahme.

Lösungen:

IBM Spectrum Scale und Spectrum Protect

Erfolge:

- Solides Speichersystem, das sich stets erweitern lässt
- Hochperformantes Storage für Forschungszwecke
- Hohe Datentransparenz
- Ideales Preis-Leistungs-Verhältnis
- Fundierte Daten zur Bekämpfung des Klimawandels



Bildmontage. Autoren: Sentinel & „Nasa Blue Marble“; Rama & NASA/ GSFC/ NOAA/ USGS

Was kann man gegen den Klimawandel tun? Wie kann man ihn überhaupt begreifen und fassen? Das Earth Observation Data Center (EODC) beantwortet diese Fragen folgendermaßen: Satellitendaten und Bodenbeobachtungen sammeln und speichern, langfristig vergleichen und analysieren – Big Data für die Klima- und Umweltforschung. EDV-Design hat sich gemeinsam mit der TU Wien und dem EODC dieser technischen Herausforderung gestellt.

Vision und Ausgangslage

Das große Ziel: Ewig verfügbare Daten

Das EODC ist ein Erdbeobachtungsdatenzentrum, in dem eine mächtige, kollaborative IT-Infrastruktur zu finden ist, die die Archivierung, Prozessierung und Verteilung von Sentinel-Satellitendaten möglich macht. Das große Ziel dabei: Diese Daten zu speichern, um sie noch in ferner Zukunft der Klimaforschung zur Verfügung stellen zu können. Den Klimawandel beobachten, bedeutet in diesem Fall: Satelliten umkreisen die Erde und machen Aufnahmen, die Aufschluss über die Erdveränderungen geben. Wie sah der Gletscher noch vor einem Jahr aus? Wie steht es um den Meeresspiegel an Stelle X? Die Gebiete werden regelmäßig von den Satelliten erfasst, um mit den Aufnahmen genau diese Fragen beantworten und Prognosen aufstellen zu können.

Je mehr Informationen vorliegen, umso granularer können die Analysen und Auswertungen für die Klima- und Umweltforschung erfolgen. Dieser nie endenden Erdbeobachtung stand auf technischer Ebene folgende Herausforderung gegenüber: Die bereits bestehenden Archive verfügten nicht über genügend Speicherplatz, um dem großen Ziel des EODC gerecht zu werden. Wie und wo kann man diese endlose Datenmasse speichern – ohne sie überschreiben zu müssen und stets auf sie zugreifen zu können?



Herausforderung

Ein Speicher ohne Grenzen

Das EODC benötigte von Anfang an ein immens großes Speichersystem, das sich dynamisch und ohne Unterbrechung vergrößern lässt. Neben der Anforderung, ein großes Datenvolumen in nur einem System unterzubringen, musste sich EDV-Design auch um die hochperformante Anbindung an den Supercomputer VSC (Vienna Scientific Cluster) kümmern. „Zudem wünschten wir uns eine moderne Architektur mit kurzen Datenwegen – im Idealfall alles in einem Gebäude – um die schnelle und ständige Verfügbarkeit der Rechenressourcen zu gewährleisten. Es sollte keine Grenzen geben, aber trotzdem ein gutes Preis-Leistungs-Verhältnis bestehen“, erklärt Wolfgang Wagner, Leiter des Departments für Geodäsie und Geoinformation an der TU Wien. EDV-Design ist langjähriger IBM Business Partner im Bereich Storage und auf High Performance Computing sowie Hochverfügbarkeit spezialisiert.

Umsetzung

Generalunternehmen EDV-Design – von Planung bis Policy

2014 begann die Vorarbeit für das grenzenlose Speichersystem in Form von Präsentationen, Diskussionen, Systemvorstellungen sowie konkreten Planungen und Konzepten. Mittelfristig ist geplant, die Daten von mehreren europäischen Satelliten zu speichern, was täglich eine Speicherkapazität von ca. 15 Terrabyte erfordern wird. EDV-Design fungierte als Generalunternehmen, das alle Phasen des Projektes betreute – von Planung und Präsentation über Aufbau und Adaption der Infrastruktur bis zur Installation und Anbindung an den VSC, Konfiguration und Inbetriebnahme. Des Weiteren erarbeitete EDV-Design gemeinsam mit dem Kunden das Layout und stellte Richtlinien zur Nutzung des Storage-Systems auf.

Lösung

Ein gigantisches Storage-System ergänzt den Supercomputer VSC

In einem ersten Schritt baute EDV-Design an der TU Wien eine Speicherkapazität von zwei Petabyte für Satellitendaten von der gesamten Erdoberfläche auf und band diese an den VSC an. Die Infrastruktur des EODC besteht insgesamt aus drei Komponenten: einer Cloud-Umgebung (OpenStack), einer Prozessierungseinheit (derzeit VSC-3) und dem Storage für das Speichern und Archivieren der Rohdaten und der prozessierten Daten. „Wir empfahlen aus technischen und preislichen Gründen, Tier-0 und Tier-1 zusammenzulegen, um somit ein administrativ vereinfachtes Gesamtsystem umzusetzen“, erläutert Martin Vlček, Prokurist bei EDV-Design. Tier-0 und Tier-1 sind mit einer Gesamtkapazität von 2 Petabyte netto mittels ESS-System aufgebaut. Diese sind sowohl über mehrfache 10 Gigabit Ethernet-Connections mit der EODC-OpenStack-Cloud-Umgebung als auch über ein Highspeed-Netzwerk mit mehreren Infiniband-Connections zum VSC-3-Cluster der TU Wien verbunden. Das ESS-System schafft einen Datendurchsatz von bis zu 26 GB/Sekunde (12 FDR Infiniband). „Als zweites bzw. drittes Tier in dieser Storage-Lösung verwenden wir eine Enterprise Tape Library. Das gesamte System verwendet IBM Spectrum Scale und Spectrum Protect für das automatisierte Information Lifecycle Management. Aktuell besteht eine Gesamt-Nettokapazität von 12 Petabyte“, führt Vlček weiter aus.

Ein gemischtes Speichersystem aus Disc und Tape war die beste und preiswerteste Lösung für das EODC. Das System ist durch und durch transparent für den Kunden, unter anderem durch die gemeinsam mit EDV-Design aufgestellte Policy zur Datensichtbarkeit: Daten, die älter als ein Jahr sind, werden auf kostengünstigere Tapes ausgelagert, damit die neuen Satelliteninformationen auf der leistungsstarken Disc Platz haben. Die auf Tape gespeicherten Daten sind stets verfügbar und absolut übersichtlich für den Benutzer – er kann so bei speziellen Projekten ganz einfach auf historische Daten zugreifen. „Wir haben hier einen soliden Storage-Grundstein erschaffen, der immer wieder erweitert werden kann. An der Grundstruktur ändert sich nichts mehr“, erläutert Martin Vlček.

„Wir haben hier einen soliden Storage-Grundstein erschaffen, der immer wieder erweitert werden kann.“

Martin Vlček, Prokurist bei EDV-Design

Erfolge und Ausblick

Solides System für Forschungszwecke

„Wir sind sehr zufrieden mit der Arbeit von EDV-Design, die ein zuverlässiger und motivierter Dienstleister bei der Umsetzung war“, sagt Wolfgang Wagner. Das überaus performante Storage-System und der schnelle Support seien dabei die Hauptvorteile. Das System wird umfassend angenommen und erfolgreich von Universitäten, Unternehmen und Forschungseinrichtungen genutzt, mit wachsenden Anwenderzahlen. Insgesamt 20 Kooperationspartner des EODC auf internationaler Ebene arbeiten bereits mit den Sentinel-Daten und profitieren dabei von dem schnellen Storage.

„Es ist weltweit einmalig, ein riesiges Erddatenarchiv unmittelbar an enorme Rechenleistung gekoppelt zur Verfügung zu haben.“

Robert Kibler, Geschäftsführer von EDV-Design

Mit den oftmals verwendeten, vergleichsweise teuren Cloud-Rechenkapazitäten kann über begrenzte Datenströme niemals so schnell auf die Datenquellen zugegriffen werden wie das im Arsenal in Wien, in lokaler Kombination möglich ist“, sagt Robert Kibler, Geschäftsführer von EDV-Design über den Erfolg. Wolfgang Wagner ergänzt mit Blick in die Zukunft: „Wir hoffen, durch die Daten und Infrastruktur des EODCs in Zukunft eine bedeutende Rolle für internationale Initiativen spielen zu können. Am Ende geht es darum, anhand fundierter Daten rechtzeitig einzugreifen, um im Rahmen des Klimawandels die Lebensqualität zu erhöhen. Und das ist eine nie endende Aufgabe, der wir durch das leistungsstarke Storage-System gewachsen sind.“

Über EDV-Design Informationstechnologie GmbH

EDV-Design wurde 1997 von Ing. Robert Kibler in Wien gegründet und ist mit einem kleinen, hochspezialisierten Team sehr erfolgreich im Bereich High Performance und High Availability tätig. Als langjähriger Enterprise-Partner führender Hersteller wie IBM, Lenovo und HPE planen, integrieren und betreiben wir die unterschiedlichsten Server-, San-, und Backup-Struktur-Lösungen für Unternehmen, Universitäten und Behörden. Mittels Konsolidierung und Virtualisierung können wir die Flexibilität, Zuverlässigkeit und Effizienz der IT-Infrastruktur wesentlich steigern. Mit „CloudiA.at“ bieten wir neueste Cloud-Technologien im lokalen Rechenzentrum.