

Case Study

PoINT Archival Gateway **EMBL-EBI**



EMBL-EBI



Das EMBL European Bioinformatics Institute (EMBL-EBI) speichert Forschungsdaten und komprimierte Daten von unter 1MB bis zu 100GB Größe, bei einem aktuellen Datenbestand von ca. 50PB. Zur Datensicherung für Backup und Langzeitarchivierung werden die Daten auf Tape geschrieben. Eine hausinterne Lösung schrieb die Daten vom Objektspeicher in ein Filesystem auf Festplatte, und dann auf Tape Medien. Dieser Ansatz erwies sich aber für die

wachsenden Datenmengen als nicht ausreichend performant. Außerdem wurde der Speicherplatz auf den Tapes nur zu 90% genutzt. Mit der Einführung von PoINT Archival Gateway verfügt das Forschungsinstitut nun über eine performante und kosteneffiziente Lösung, um die Daten über die standardisierte S3 Schnittstelle direkt auf Tape zu schreiben und dem Datenwachstum begegnen zu können.

Herausforderung

- Backup und Langzeitarchivierung von Objektdaten auf Tape
- Skalierbarkeit für stetig wachsende Datenbestände
- Hohe Anforderungen an Performance und Kapazität

Lösung - PoINT Archival Gateway

- Standardisierte S3 Schnittstelle statt S3 Glacier Kommandos
- Native Tape-Anbindung
- Node-basierte Software-Lösung für hohe Skalierbarkeit und Flexibilität

Vorteile

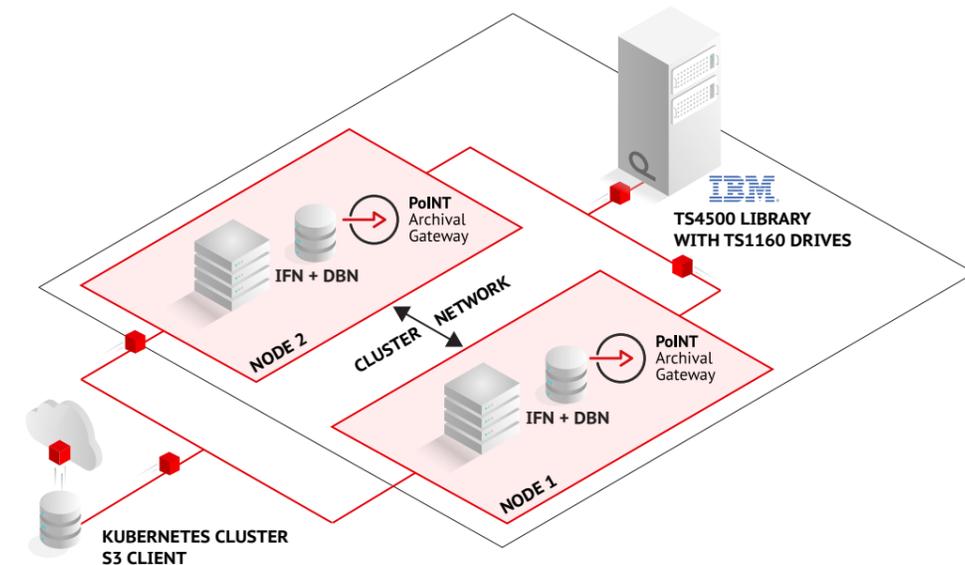
- Kosteneffizienz durch sparsamen Hardware-Einsatz
- Performante und flexible Speicherlösung für Backup und Archivierung
- S3 Datenstrom ohne Festplatten-Buffer direkt auf Tape
- Hardware- und herstellerunabhängig

Über das EMBL European Bioinformatics Institute (EMBL-EBI)

Das EMBL European Bioinformatics Institute (EMBL-EBI) ist weltweit führend bei der Speicherung, Analyse und Verbreitung großer Datensammlungen aus der biologischen Forschung. EMBL-EBI hilft Wissenschaftlern dabei, die Möglichkeiten dieser Datenmengen zu nutzen. Es fördert ihre Fähigkeiten, komplexe Informationen auszuwerten und Entdeckungen zu machen, die der Menschheit zugutekommen. Das Institut ist richtungsweisend in der computergestützten biologischen Forschung. Die Arbeitsfelder umfassen

Methoden der Sequenzanalyse, die multidimensionale statistische Analyse und die datengesteuerte biologische Forschung, von der Pflanzenbiologie bis zur Entwicklung von Säugetieren und zur Erforschung von Krankheiten.

EMBL-EBI ist Teil des European Molecular Biology Laboratory (EMBL) und befindet sich auf dem Wellcome Genome Campus, der weltweit bedeutendes wissenschaftliches und technisches Know-how im Bereich der Genomik bündelt.



Das EMBL European Bioinformatics Institute (EMBL-EBI) nutzt festplattenbasierte Objektspeichersysteme als Primärspeicher für die Speicherung von Forschungsdaten aus Genomik-Projekten. Um die Daten langfristig zu speichern und bei Ausfällen der Primärspeichersysteme schnell und zuverlässig auf die Daten zugreifen zu können, wird eine vollständige Kopie auf Tape gesichert. Für diese Replikation auf Tape wurde ein hausinternes Skript genutzt, das aber nicht die notwendige Performance lieferte und den Speicherplatz auf den Tapes nur zu 90% ausnutzte. Angesichts des bereits vorhandenen Datenvolumens von 50PB und eines erwarteten Wachstums von mehr als 1PB pro Monat suchte man eine Alternative, die die Anforderungen hinsichtlich Performance, Kosten und Zukunftssicherheit erfüllte.

Mit PoINT Archival Gateway hat das Institut eine Lösung eingeführt, die die Objekte als Objekte direkt über die standardisierte S3 Schnittstelle auf Tape schreibt. Als Zielspeicher wird initial eine IBM TS4500 Library mit sechs IBM TS1160 Laufwerken eingesetzt. PoINT Archival Gateway benötigte dafür lediglich einen 2U Server. Im Zuge der aktuellen zweiten Projektphase wurde die Installation auf 12 Laufwerke und zwei 2U Server erweitert.

Entscheidend für die Auswahl der PoINT Software war zum einen, dass PoINT Archival Gateway die Standard-

S3-Schnittstelle nutzt. Es sind keine umständlichen S3 Glacier Kommandos notwendig. Zum anderen benötigt PoINT Archival Gateway keinen Festplatten-Buffer, um auf Tape zu schreiben. Die Nutzung des RAM Cache ermöglicht hohe Performance-Werte und erhebliche Einsparungen im Festplattenbereich.

PoINT Archival Gateway benötigt nur einen 2U Server für 6 Laufwerke. Durch den sparsamen Hardware-Einsatz erweist sich die Lösung als besonders kosteneffizient und gleichzeitig leistungsstark.

„Wir sind sehr zufrieden mit dem Einsatz von PoINT Archival Gateway und besonders auch mit der Zusammenarbeit und dem Austausch mit PoINT und dem Partner Mercia“, so Marc Riera (Software Development and Operations – Service and Data Management Technical Coordinator beim EMBL-EBI). „Auf das künftige Datenwachstum fühlen wir uns gut vorbereitet und freuen uns, die weiteren Entwicklungsschritte von PoINT Archival Gateway mitzugehen.“

So liefert PoINT Archival Gateway zum Beispiel auch für die geplante Tape-Migration die geforderte Lese- und Schreibrate von mindestens 1PB pro Woche. Speicherkapazität und Performance sind bedarfsgerecht

skalierbar. Die Herstellerunabhängigkeit der Software-Lösung ermöglicht es, die Hardware unterschiedlicher Hersteller und unterschiedliche Laufwerksgenerationen flexibel zu kombinieren. So sind Zukunftssicherheit und ein hoher Investitionsschutz gegeben.

About Mercia Solutions

Mercia Solutions based in Castle Donington, England. The british reseller have years of experience in data management and data lifecycle solutions. They can help you from planning through to complete integration of PoINT Software. Mercia Solutions specialise in supporting resellers and channel partners. The reseller has demonstration equipment and pre-sales support available to assist you. Get in touch today and let them help you with your projects.

Über PoINT

PoINT Software & Systems GmbH ist spezialisiert auf die Entwicklung von Softwareprodukten und Systemlösungen zur Speicherung und Verwaltung von Daten auf allen verfügbaren Massenspeichern, wie optischen Medien, Festplatten, Magnetbändern, Cloud und Object Store. Die enge Zusammenarbeit mit seinen Technologie-Partnern ermöglicht dabei die frühzeitige Unterstützung neuer Speichertechnologien.

Hohe Flexibilität, Berücksichtigung des Workflows und regelbasiertes Data Management erlauben eine effiziente Nutzung der Hardware und helfen, durch das Datenwachstum entstehende Kosten und Probleme zu reduzieren.

Weitere Informationen und eine Evaluierungsversion sind unter www.point.de verfügbar.