European Bioinformatics Institute Drives Innovation with Delphix

Supports critical life science research with faster data

EMBL-EBI

INDUSTRY
Bioinformatics research and services

HEADQUARTERS
Cambridge, UK

EMPLOYEES
570

ABOUT EMBL-EBI

The European Bioinformatics Institute (EMBL-EBI) is a global leader in the storage, analysis, and distribution of life sciences datasets, helping scientists exploit complex information to make discoveries that benefit mankind. Over two million researchers access their freely available life science data each year. EMBL-EBI is a non-profit, intergovernmental research organization based in the UK, funded by 21 member states and two associate member states.

BACKGROUND

EMBL-EBI manages well over 50 Petabytes of data, with this amount doubling every year. Researchers in medicine, agriculture, and environmental science issue over 12 million requests per month for this freely available life science data, which is managed jointly with collaborators in the US and Japan. Data from genome sequencing uses most of the storage available at EMBL-EBI, and demand in this area of science is growing rapidly as the price of the supporting technology continues to fall. Scientists at EMBL-EBI regularly add information about genome sequencing and other data types, and need to find innovative ways to improve database efficiency and scalability.

CHALLENGE

The collection, curation, and release of reference genome data is vital for research activities worldwide – especially in the area of personalized medicine which will be a major driver for healthcare innovation. However, the sheer size and complexity of the data makes it increasingly difficult to move both internally and externally.

"Reputation is the currency of research. Delphix Data as a Service allows us to release our data faster and more frequently than ever before reducing the time to release and deploy our data into production making most effective use of our restricted resource."

STEVEN NEWHOUSE
Head of Technical Services, EMBL-EBI

DELPHIX

It took EMBL-EBI up to three months to prepare a data release. Much of this time was spent passing copies of databases from one team to another, adding extra information about different molecules and interactions along the way. The 12 million monthly requests also involved a time-consuming process of repeatedly copying and migrating datasets from the development and analysis data center in Hinxton to the public services data center in London.

A third data center at an undisclosed location is used for disaster recovery and all databases and files replicated regularly across all three locations. With individual datasets as big as seven Petabytes, its metadata needs to be stored in over 500 repositories across Oracle, MySQL, PostgreSQL and NoSQL databases.

EMBL-EBI needed a platform that could handle very large volumes of data sitting across multiple database sources, and provide full copies across multiple locations. Developing and testing genome-sequencing data requires data at scale, meaning data subsetting was not an option. Because of EMBL-EBI's distributed infrastructure, database release and replication were vital.

SOLUTION

EMBL-EBI started a data agility project based on Delphix with the goal being to virtualize the databases so operations teams can prepare and release research data faster and more frequently than before. EMBL-EBI successfully deployed Delphix and now hosts over 50 virtual database environments supporting test and development operations. It also provides additional read-only copies of production databases for general internal use while further production operations continue.

> *"This kind of agility was just not possible without Delphix."*
> **MANUELA MENCHI**
> **Database Team, EMBL-EBI**

RESULTS

Developers and engineers can now self serve their own temporary data environment on demand and in minutes. They are also able to rewind data to any point in time without needing to tap into archives to retrieve historical data. EMBL-EMI projected that Delphix will:

· Reduce data preparation timeframes by 20%

· Increase exploratory work, benchmarking or development activity

· Increase output without additional development, curation or DBA staff

· Reduce the total database storage footprint by 70%

· Deliver more frequent releases of data

Future plans for Delphix include increased consolidation of data infrastructure, replication between data centers, and backup and recovery enhancements.

KEY BENEFITS

20%
faster preparation of datasets

Increased
frequency of releases

Reduced database storage by
75%

Improved
service while maintaining flat headcount

ABOUT DELPHIX

Delphix's mission is to free companies from data friction and accelerate innovation. Fortune 100 companies use the Delphix Dynamic Data Platform to connect, virtualize, secure and manage data in the cloud and in on-premise environments. For more information visit www.delphix.com.

DELPHIX