

Success Story: Cloud Native Computing Enables  
Scientific Knowledge Management Over All Relevant  
Information Since the Inception of Time



#### **Customer**

An international company

#### **Partner**

PRODYNA SE

#### **Products and Services**

Azure Kubernetes Service and many others

#### **Customer**

The customer is one of the world's largest enterprises that since its inception has continually invested a considerable portion of its financial resources into research and development.

#### **Challenge**

The production of its own publications, patents, and research reports together with scientific publications has over time led to the creation of a heterogeneous landscape of multiple „big data“ knowledge systems. Research scientists working on the discovery of new materials and

and processes require the availability of the entire company knowledge for search and comparison within a single system. A typical detailed search and analysis would take significant time and effort.

The software architecture for the actual product required careful consideration, in particular due to the non-functional requirements. As the system was to be capable of answering requests in the scope of the entire company knowledge since the beginning of time, the system had to be capable of handling enormous amounts of data. The system was required to be able to ingest large amounts of new data at regular intervals and subsequently process the new data in the context of existing data and write the reprocessed results without a significant performance loss. Parts of the data are confidential and required an appropriate security concept.

From the initial analysis onwards, it was clear that the system would have high dynamic requirements on computing resources since data processing has a 10-20 times higher compute requirement than normal search request operations. Data delivery and reprocessing the documents would occur only every few days, making dynamic scaling mandatory. It was obvious this system would benefit from using a cloud provider with a pay-as-you-use model if it were to be financially viable.

Due to the combination of available security concepts and services, one particular example being Azure Information Protection, the Microsoft Azure Cloud was selected to host the final application.

**About PRODYNA:** PRODYNA is an innovative IT consultancy specializing in the creation of custom software solutions and serving the needs of corporate enterprises across the European continent. PRODYNA is a Microsoft Partner, Kubernetes Certified Service Provider, Kubernetes Training Partner, and a member of the Cloud Native Computing Foundation.

## Result

The application is built using the current software architectural principles and cloud native technologies. The central knowledge graph part of the application is based on the Neo4j graph database which is accessed through a microservice based architecture created with Spring Boot. Docker is used to package the application into container images which are subsequently executed on a Kubernetes cluster using the Azure Kubernetes Service. A GitOps approach is used for continuous integration and Kubernetes cluster management using Terraform as the technology for describing the entire infrastructure as code. Terraform scripts are executed under the full control of the customer with human approval steps. Further notable components of the application are Azure Blob Storage for the storage of binary objects, Elastic Search as the engine behind the full-text index and ChemAxon tools for the molecular substructure search.

The running application is monitored using the metric analysis and visualization suites, Grafana and Prometheus. During normal operations, ten Kubernetes cluster nodes are required for running the core databases, search operations, and user interface for end user requests. Following the delivery of new data, the application is automatically scaled out to approximately 150 nodes for data processing. Under these

conditions, the required computing resources extend to around 2300 CPU cores and 16 terabytes of RAM.

The massive amounts of data needing to be handled and the extremely dynamic computing resource requirements precluded such an application being hosted at the customer premises. Although much of the data being processed is classified as confidential, from the standpoints of time-to-market, and financial viability, this system could only be hosted in a cloud environment with pay-per-use temporary resources. The entire environment was security audited by an external IT-security company and combined with Azure Information Protection meaning that all downloaded documents are encrypted and can be viewed or edited by company employees following a two-factor authentication process.

As such, this project is a prime example of cloud computing enabling systems and goals that would otherwise be unfeasible and is a great blueprint for following cloud native projects.

## Technologies:

- Azure Kubernetes Service
- Azure Information Protection
- Azure Storages (Blob)
- Azure Container Repository
- Azure Firewall
- Azure Vault
- Azure Log Analytics
- Neo4j Graph database
- Spring
- Terraform
- Docker
- Grafana
- Prometheus
- ElasticSearch
- RabbitMQ