SAP Innovation Awards 2019 Entry Pitch Deck

Breakthrough in protein analysis advances war on cancer

Technical University of Munich – TUM

THE BEST RUN **SAP**

Video: SAP HANA for Proteomics for Personalized Medicine

https://www.youtube.com/watch?v=ao4oStycKnw
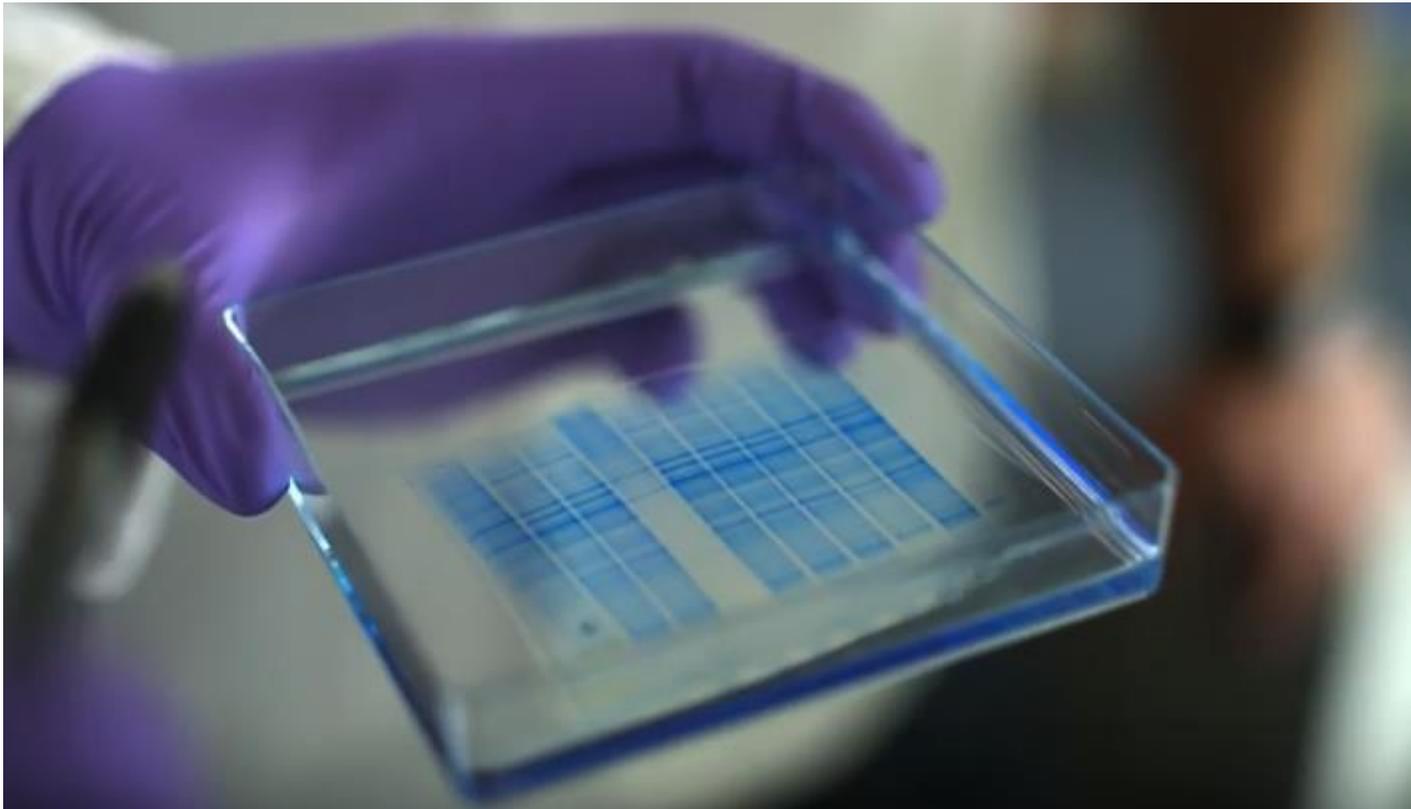
Breakthrough in protein analysis advances war on cancer
Technical University of Munich

"Quote"

"With our ProteomicsDB based on SAP HANA running on IBM Power Systems, we have the perfect platform to help medical scientists around the world to create more-effective treatment options – to save lives in the future."

Prof. Dr. Bernhard Küster

Head of the Chair of Proteomics and Bioanalytics

TUM School of Life Sciences Weihenstephan, Technical University of Munich

Challenge

Map the entirety of proteins in the human body to create the proteome – and make this possible by finding a database solution fast, flexible and powerful enough for this endeavor

Solution

Create and continually populate the ProteomicsDB – powered by SAP HANA on IBM Power system – and integrate it seamlessly with statistics and analytics applications running on an R-server

Outcome

Revolutionized the speed and depth of understanding about the relationships between proteins and how they interact with drugs, enabling more effective and targeted treatments of cancer patients.

>1,000

published academic research papers reference ProteomicsDB

243

cancer drugs analyzed to enable personalized treatment in the future

80%

or 15721 of 19629 proteins of the human proteome quantified and made accessible online
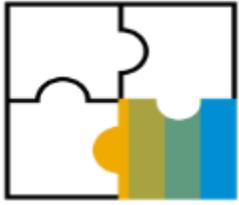
Partner Information

IBM
Provided the Power Systems server for outstanding performance and high availability plus IBM Storwize data storage systems and IBM Spectrum Virtualize software for streamlined storage management

"SAP HANA on IBM Power Systems is the ideal infrastructure to build our protein-centric database – helping scientists make the most of large collections of proteomics data. IBM Power provides us with the performance, high availability, straightforward maintenance and continuity we need to ensure the sustained operation of the ProteomicsDB."

Dr. Harald Kienegger, Managing Director of the University Competence Center (UCC), Technical University of Munich

"The ProteomicsDB is definitely faster since we moved to SAP HANA on IBM Power Systems. The rich user interface enables scientists to tap into a wealth of data and gain new insights by adjusting different parameters and evaluating the changes immediately."

Dr. Mathias Wilhelm, Group Leader Bioinformatics, Chair of Proteomics and Bioanalytics, TUM School of Life Sciences Weihenstephan, Technical University of Munich
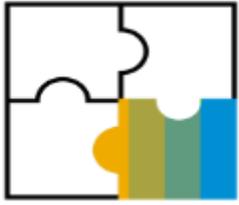
Business Challenge & Objectives

To uncover new insights into biochemical processes within the human body, researchers need to explore and analyze huge volumes of experimental data 'on the fly.' This requires complex data aggregations that cannot be handled by conventional databases as they are far too slow and cumbersome for such demanding research tasks. To overcome such limitations, TUM was looking for a high-performance database solution that would support a simple, flexible and efficient approach to rapid data analysis.

To create a powerful platform to store, analyze, and visualize proteomics data without constraints, the Chair of Proteomics and Bioanalytics at TUM worked closely with SAP to build ProteomicsDB, leveraging the capabilities of the SAP HANA business data platform. The system also needed to scale continuously to accommodate growing data volumes. To achieve such speed, flexibility, and scalability, the ProteomicsDB team enlisted help from IBM in migrating ProteomicsDB to the latest release of SAP HANA and moving the database to a newly installed, dedicated IBM Power Systems server.

Project / Use Case Details

In the treatment of complex molecular diseases such as cancer, one size does not fit all – because the presence and abundance of proteins differ from person to person. Thus, decoding human proteins is key to understanding these diseases and developing personalized medication. The challenge is in mapping out all proteins in the body, and creating the 'human proteome' in the same way that researchers have mapped out the human genome.

In 2013, a team of 30 experts in cell biology, chemistry, biochemistry, mass spectrometry, and bioinformatics thus joined forces with SAP to create the ProteomicsDB which went live in 2014. This massive in-memory database uses SAP HANA for data storing and native modeling. SAP HANA allows the team to implement a simple database design that eliminates precomputed data aggregations. In 2017, ProteomicsDB was migrated to the latest release of SAP HANA and moved to a newly installed, dedicated IBM Power Systems server to optimize operations and accommodate growing data volumes. The transition was planned by TUM's University Competence Center (UCC) organization, which is also hosting the ProteomicsDB. The highly automated infrastructure management based on IBM Power Systems enables TUM to smoothly operate SAP HANA business data platform for the ProteomicsDB.

Today, 8.85 TB of proteome data are at the disposal of researchers to rapidly collect and analyze complex data sets, to help them uncover new insights from existing experimental results, to advance their understanding of biochemical processes and to advise clinicians on personalized treatments thus enabling truly evidence-based medicine.

ProteomicsDB is instrumental in driving major breakthroughs in cancer research and treatment – it has been referenced in more than 1,000 scientific articles, among them a publication in the journal NATURE.

Benefits and Outcomes

Business / Social

- Provide advice to leading doctors on targeted, personalized treatments in a 'Molecular Tumor Board'
- Researchers in science, healthcare and the pharmaceutical industry gain a better understanding of the mechanisms associated with cancer drugs in patients
- The shift from anecdotal evidence (and guesswork) to hard evidence supports organizations in focusing their R&D effort on outcomes
- Solution can be commercialized and offered as a service for use by other research teams that e.g. look into proteomics of plants and animals – thus further adding to the understanding of proteomics

IT

- Faster response to research inquiries based on enhanced system performance
- Robust, versatile, and high-speed processing capabilities that support several types of complex, multidimensional data and the development of a variety of features
- SAP HANA multi-tenant database containers support the efficient use of compute resources
- IBM solutions for storage technology and virtualization provide the flexibility to expand capacity easily and rapidly when needed

Human Empowerment

- Without such a high performance database as SAP HANA, running on IBM Power Systems, the research endeavor wouldn't have been possible at all
- Fast queries across vast amounts of data enable users to identify previously unknown correlations that make predictions of drug sensitivity possible
- ProteomicsDB features a public web interface, enabling external researchers to contribute to and benefit from this unique database solution

Architecture

ProteomicsDB consists of three major layers.

- The bottom layer is the data layer, providing information to the calculation layer. It consists of seven major modules enabling the storage and retrieval of meta data, annotations and quantitative information associated with proteins and biological systems.

- Due to in-memory storage of the data layer using SAP HANA, calculations using the calculation engine (structured query language), graph engine and other integrated programming languages (e.g. R and Python) are highly efficient.

- The results of these calculations can be explored in the presentation layer, offering a variety of different interactive visualizations via the web interface or systematic access via the ProteomicsDB application programming interface (API).

Source: ProteomicsDB, Nucleic Acids Research, 2018, Vol. 46, Database issue D1271–D1281
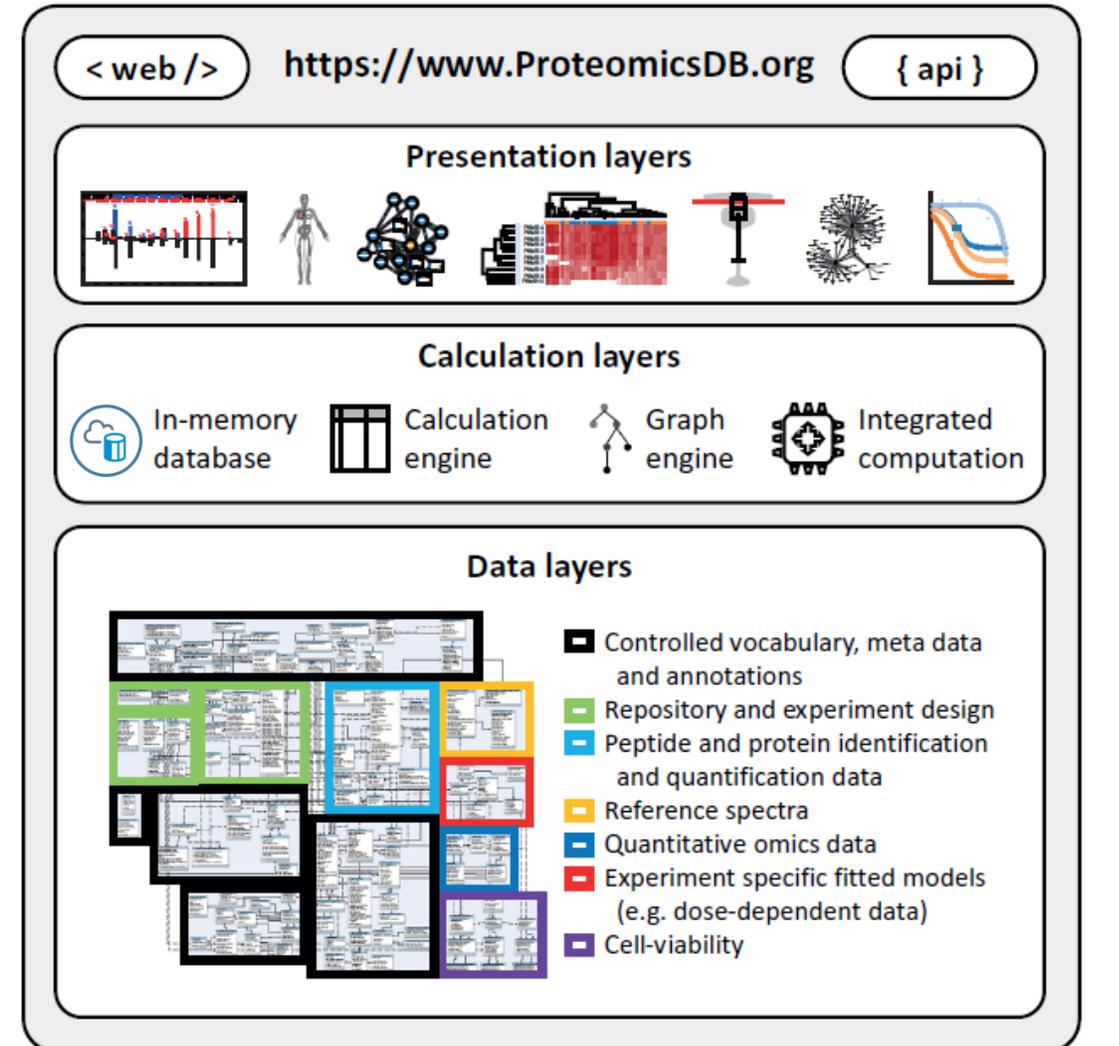
Deployment

Date of Deployment:

Deployed in 2014, and moved to new IBM Power Infrastructure in 2018, publicly available: https://www.proteomicsdb.org/

Number of live users:

800 registered + varying number of ad-hoc users

SAP Technologies Used: Live?

SAP HANA business data platform	✔
SAP Predictive Analysis Library	PoC

ProteomicsDB was initially hosted at SAP. In June 2017, it was migrated to TUM's SAP University Competence Center (UCC) on IBM Power. 2018 saw a hardware refresh plus a scale-up from 2 TB to 6 TB of memory.

Server Processor: IBM Power Systems™ E870C

Linux Distribution: SUSE Linux Enterprise Server for SAP Applications

Emerging Technologies and Use Cases

The following Emerging Technologies and use-cases are part of the project and describe the contribution

	Technology or Use Case	Yes/No	Contribution to Project
1.	Machine Learning / Artificial Intelligence	Yes	TUM will leverage Machine Learning (ML) capabilities by using the predictive analysis library in SAP HANA. Moving ML workloads directly into SAP HANA will help to reduce overhead and to streamline and accelerate data processing.
2.	IoT	No	
3.	3D printing	No	
4.	Blockchain	No	
5.	API Economy / Integrate the Intelligent Enterprise	No	
6.	Cloud Native / Event Based Architectures	No	
7.	Extending the digital core with SAP CP / ABAP in SAP CP	No	
8.	SAP Leonardo Application (extending SAP application, using Industry Innovation Kits or result of Design Thinking workshop)	No	