PaSER 2024: a future-proofed renovation for better robustness, ease of use, flexibility, and scalability.

Sebastian Wehner[1], Lennard Pfennig[1], Marc-Antoine Beauvais[2], José Abuin[1], Paolo Ballesteros[2], Duy Nguyen[2], Dennis Trede[1], Christopher Adams[2], Jonathan Krieger[3], Tharan Srikumar[3] and Sven Brehmer[1]

[1]Bruker Daltonik GmbH, Fahrenheitstraße 4, 28359 Bremen, Germany [2]Bruker Scientific LLC, 40 Manning Road, Manning Park, Billerica, MA 01821, USA, [3]Bruker Ltd., 2800 Highpoint Dr. Suite#206, Milton, ON L9T 6P4, Canada

Introduction:

Parallel search engine in real-time or PaSER was developed together with the Yates lab to take advantage of GPU-powered database search. Since its initial conception, PaSER has been transforming into a proteomics data analysis platform that can integrate 3rd party tools while utilizing the concept of stream and stream-processors to realize fully customizable real-time processing workflows including on-the fly decision making based on the data generated. However, technical challenges continuously hindered forward momentum on the previous technology stack. PaSER has now been re-written to encompass 4 fundamental principles: (1) ease of use, (2) scalability, (3) flexibility, (4) robustness. The new PaSER platform utilizes an open-source, cloud native technology stack that is well supported by the community and technology giants providing a stable base for future growth. Using this platform, we can illustrate the optimal usage of system resources that are automatically balanced depending on the workload.

Concept and Goals:

PaSER 2024 is designed as a proteomics analysis platform that is flexible, scalable, robust and easy to use, for users as well as developers, making it an easily extensible\interrogatable system. It focused on these key points:

- Usability
- Elastic and Scalable processing
  - Integrate additional compute resources
  - Parallelize and distribute computing tasks if possible
  - Optimal usage of resources (including GPU)
- Flexibility
  - Let users decide how to lay out processing workflows
  - Let users decide which algorithms to use
  - Let users decide what questions they want to ask (queries)
- Deployment
  - Single workstation
  - Multiple workstation
  - Local – cloud hybrid

  } Distributed computing
- Security
- Open
  - To integrate “external” services
  - To get integrated into “external” systems

  } Distributed computing

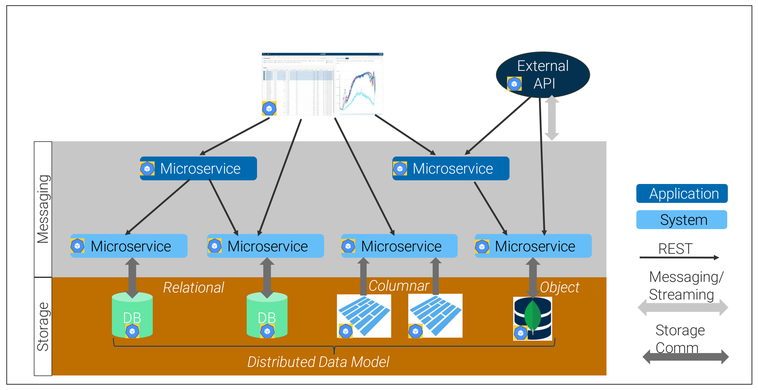


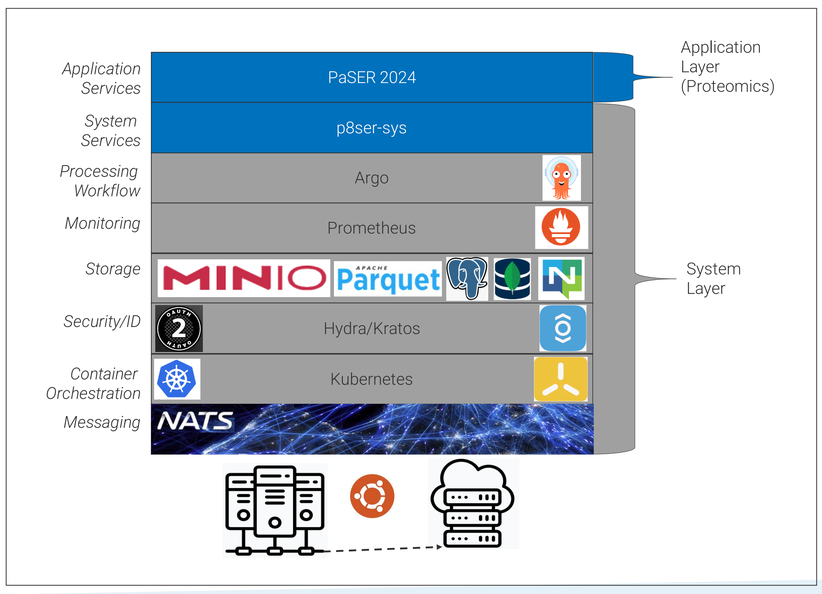
Fig. 1: Architecture overview of PaSER. PaSER is composed of application (Proteomics) level microservices and system level microservices interconnected via standard REST services, messaging and streaming and storage communication protocols.

Results:

The product of which is a microservices architecture-based system consisting of a base-system layer and application domain layer interconnected via REST protocols, streaming and storage communication (see Fig 1). The application layer contains services and workflow task stream-processors such as database search tools (e.g., ProLuCID), FDR estimation (e.g., DTASelect) and label free quantitation (e.g., Census). The system layer contains all the services and support infrastructure that are not directly “domain” specific. This provides a generalized platform that can be tweaked for specific applications. To further flexibility, each microservice or workflow-task stream-processor can access databases (specific or shared), large data pools or large object stores.

When evaluating and choosing a technology stack, in addition to our goals, we also choose to utilize tools that are well supported by a large open-source community, as well technology giants.

Fig. 2: Technology overview of PaSER. A cloud native technology stack consisting of projects endorsed by the Cloud Native Computing Foundation provides a stable foundation backed by a large community and many technology giants.



The Cloud Native Computing Foundation projects (https://www.cncf.io/) were preferentially selected as part of PaSER technology stack.

The great technology stack needed an equally elegant graphical user interface. PaSER’s frontend was designed by an expert UI/UX firm. Input and feedback on prototypes were provided by many users over the span of 9 months. The initial prototypes have been refined to provide a UI that is intuitive to use but provides all the required information to a variety of different users and use cases. Data visualization was one of the primary focuses of the re-design and now provides a vehicle from which further customization (application specific or user specific) can be made in future iterations.

Fig. 3: The redesigned graphical user interface for PaSER. A selections of screenshots highlighting different aspects of the user interface, including many visualization options that allow drilling down into the data.

Conclusion:

- PaSER has been re-designed from the ground up
- Powered by a technology stack backed by CNCF projects, with large open-source community and technology giant's support
- Has an intuitive user interface with a large focus on data visualization
- Supports many different deployment cases

Technology