View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by KITopen

Architecture to fulfill the requirements of Data Intensive Applications

M. Sutter[1], R. Stotzka[1], V. Hartmann[1], T. Jejkal[1], A. Trunov[2], J. van Wezel[2]

[1] Forschungszentrum Karlsruhe GmbH, Institute for Data Processing and Electronics, Karlsruhe, Germany
[2] Forschungszentrum Karlsruhe GmbH, Steinbuch Centre for Computing, Karlsruhe, Germany

Abstract

The high throughput and high content microscopy facility at KIT will produce data in the order of PB's per year. The data is stored and archived at the Large Scale Data Facility (LSDF) at KIT, a new storage infrastructure developed to fulfill the requirements of data intensive applications. The computational requirements for analyzing the data are covered by Grid, HPC and Cloud computing resources in combination with seamlessly integrated software environments, allowing the access to the LSDF. All components can be controlled via a simple interface: The Data Browser, allowing scientists to focus on the evaluation and processing of data within near real-time.

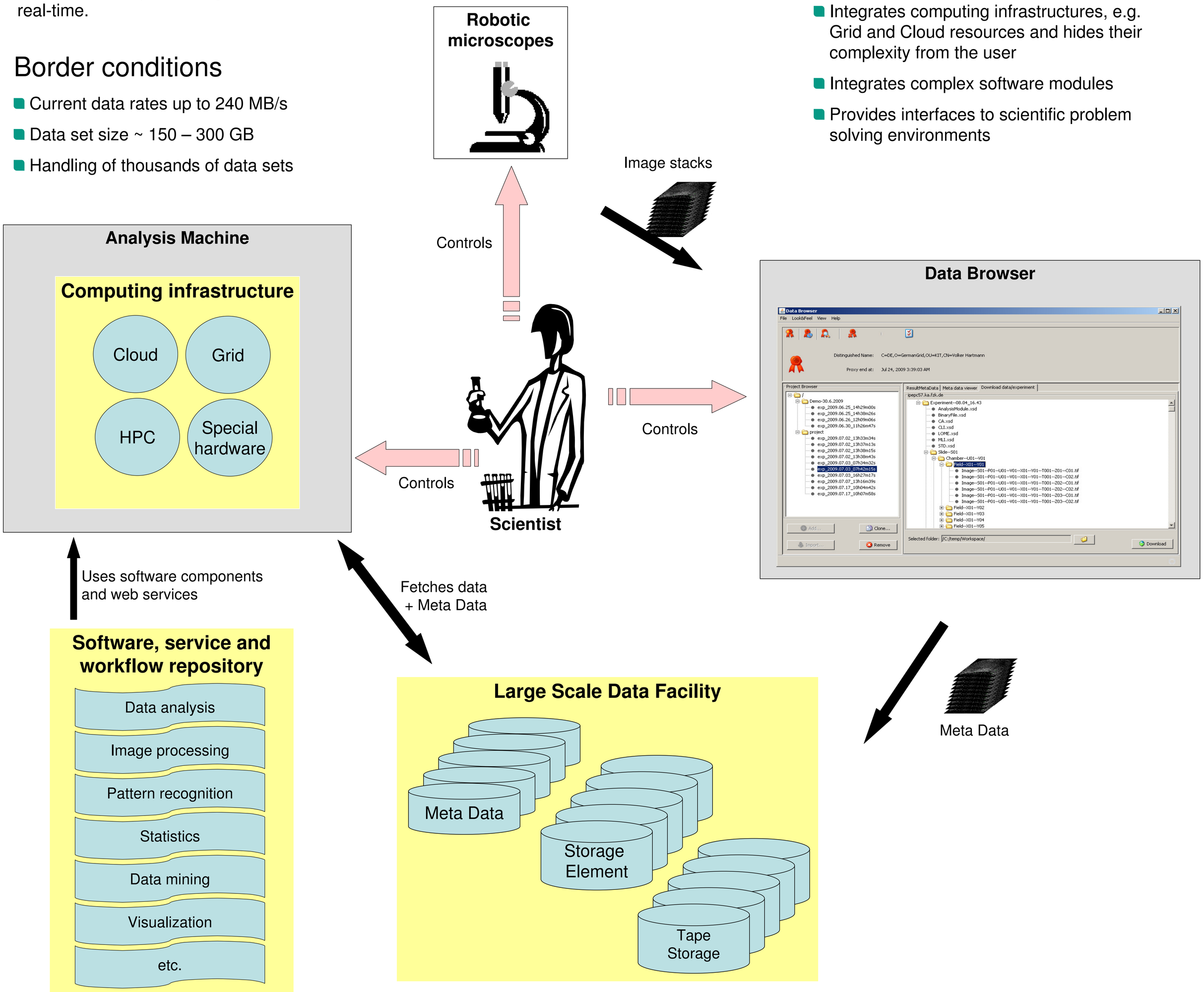
Border conditions

- Current data rates up to 240 MB/s
- Data set size ~ 150 – 300 GB
- Handling of thousands of data sets

Aims

- Persistent storage and archiving of data including Meta Data
- World-wide secure access
- Temporal and spatial analysis in 3D in near real-time
- Easy-to-use
- Easy to extend with new algorithms and workflows

Proposed architecture



Software components

Data Browser:

- Easy-to-use access to Grid storage elements within the LSDF
- Extensible GUI integrating security, data organization, visualization and access to the Analysis Machine
- Data and meta data models designed and implemented by an operative service

Analysis Machine:

- Software service environment to run interactively, to develop and to debug data analysis workflows
- Integrates computing infrastructures, e.g. Grid and Cloud resources and hides their complexity from the user
- Integrates complex software modules
- Provides interfaces to scientific problem solving environments

Contact:
Michael Sutter
Email: michael.sutter@ipe.fzk.de
Phone: +49 7247 82-5676

KIT – The Cooperation of:
Forschungszentrum Karlsruhe in der Helmholtz-Gemeinschaft
Universität Karlsruhe (TH)
Research University · founded 1825