Hasso Plattner Institut
IT Systems Engineering | Universität Potsdam

Data Management for Digital Health
Kick-off

Dr. Matthieu-P. Schapranow

Data Management for Digital Health

Summer 2017

Agenda

- Lecture organization
  - Structure
  - Schedule
  - Contents
- Facts you should know

Lecture Organization
Administrative Details Time / Dates

- Location: HPI Campus II, D-E.9/10
- Tuesdays 9.15am-10.45 am (s.t.)
- Thursdays 11.00am-12.30pm (s.t.)

- Enroll for the lecture until Apr 28, 2017 (firm deadline)
- Website and further details: https://hpi.de/plattner/teaching/summer-term-2017/data-management-for-digital-health.html

Lecture Organization
Grading

- Credit points: 6 ECTS

- Grading will be determined by the following individual parts, each part must be passed individually:
  - Intermediate exercises
  - Exam

http://www.hpi.uni-potsdam.de/fileadmin/hpi/presse/Fotos/campus_und_gebaeude/20111017_HPI_Hoersaal.jpg

Lecture Organization
What you can expect from us

- Broaden your horizons in the fields of
  - Digital Health,
  - Life sciences, as well as
  - Data challenges and opportunities
- Work with real-world data, real-world use cases
- Hands-on experiments of selected tools
- Invited talks by key experts in the field
- Get experience in collaborative project work

http://i.kinja-img.com/gawker-media/image/upload/s--cRElB5AZ--/1865smw5hbbt6jpg.jpg

Lecture Organization
What we expect from you

- Commitment to the lecture and exercises
- Attend lectures regularly
- Participate in group discussions, expert talks, and excursions
- Perform autonomous research to dig deeper into the topics
- Contribute with your expertise also to your colleagues
- Update supervisors on any issues you might encounter

http://i.kinja-img.com/gawker-media/image/upload/s--cRElB5AZ--/1865smw5hbbt6jpg.jpg

Lecture Organization
Sneak Preview

- Hands-on work with

Credit: Caine et al. *A 3D-DNA Molecule Made of PlayMais*. 2015

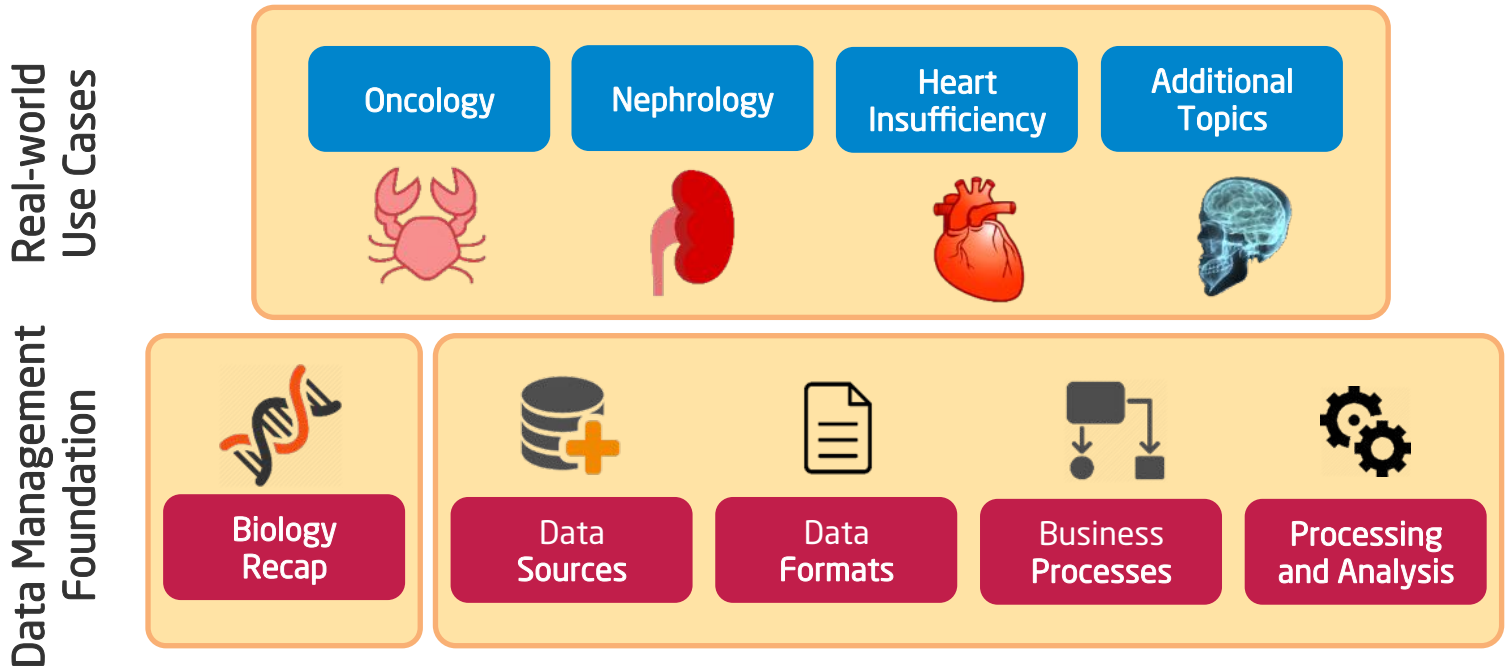
- Discover the reality

Credit: Delaware State News/Dave Chambers

- Discuss with experts

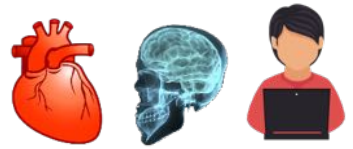
Credit: Acuitus Medical

Lecture Structure



Schedule

- Biology recap
- Lecture Kickoff
- Data sources and data formats
- Introduction
- Use Case Oncology



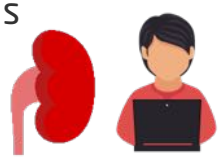
- RNAseq processing
- Unsupervised learning
- Use Case Heart Insufficiency
- Special topics

April | May | June | July

- Exam

- Expert talks
- On site visits

- Clinical prediction models
- Supervised learning
- Use Case Nephrology



- Exercises

Lecture Contents
Biology Recap

Hasso Plattner Institut

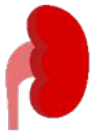
- Discovery of the Human Genome
- Components of cells
- How does DNA look like
- DNA/RNA sequencing technologies

Lecture Contents
Use Case Oncology

Hasso Plattner Institut

- Personalized, stratified, and precision medicine
- Clinical oncology process
- Identifying options for cancer treatment
- Retrieving information on cancer biomarkers
- Data formats and properties of data
- Distributed computing, process workflows (BPMN)
- Application examples: Genome Browser, Medical Knowledge Cockpit
- Expert talk oncology

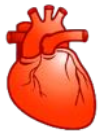
Lecture Contents
Use Case Nephrology



Hasso Plattner Institut

- Clinical prediction models
- Acute vs. chronic kidney diseases
- Supervised learning, e.g. SVM, Bayesian networks
- Developing a clinical model
- Intensive care data: MIMIC III database
- Data analysis: RapidMiner
- Expert talk nephrology

Lecture Contents
Use Case Heart Insufficiency



Hasso Plattner Institut

- Systems medicine
- Use Case Heart Failure
- Unsupervised learning: clustering techniques
- Differential expression
- Discriminate healthy and diseased hearts based on RNAseq data
- Multiple factor analysis
- Expert talk cardiology

Lecture Contents
Special Topics

Hasso Plattner Institut

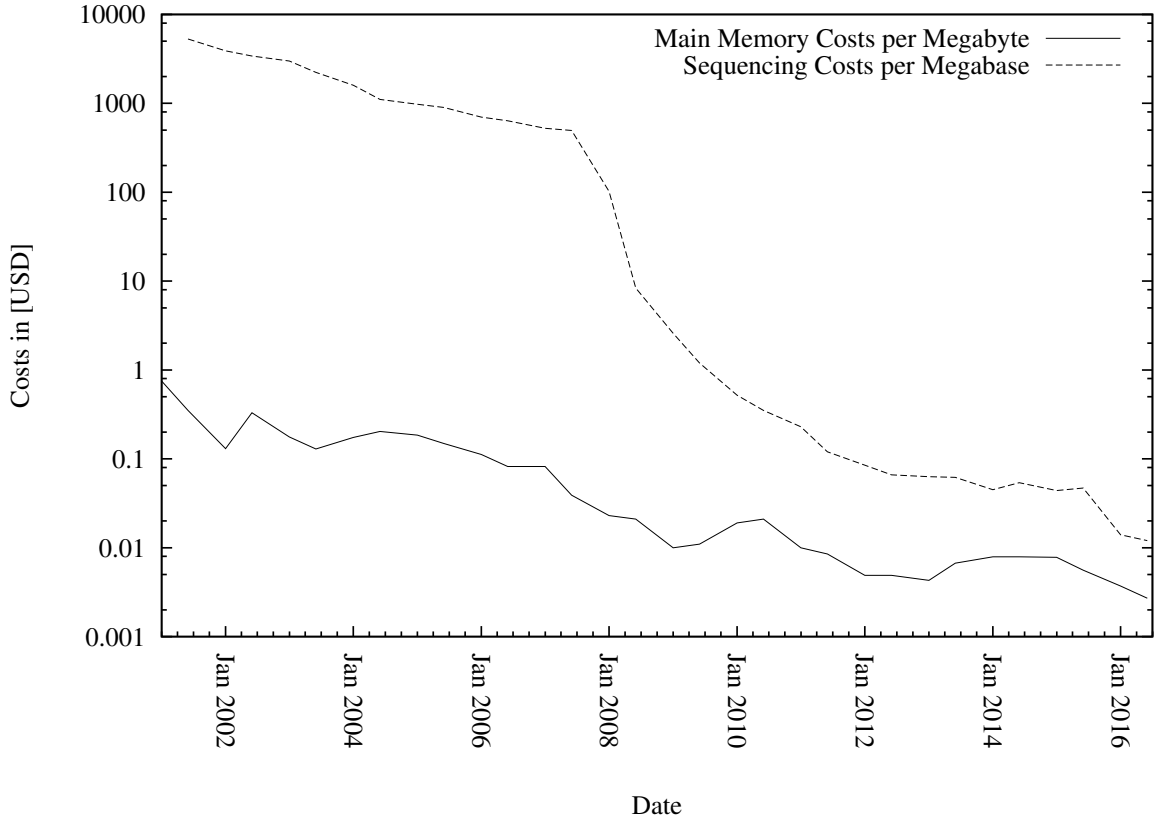
- Imaging and image data analysis
  - State-of-the-art tools
  - Building deep learning models
  - Segmentation and classification of brain CT scans
- Text analytics
  - Automated discharge letter generation
  - Template filling
  - Summary generation
  - Question answering systems

Any Questions?

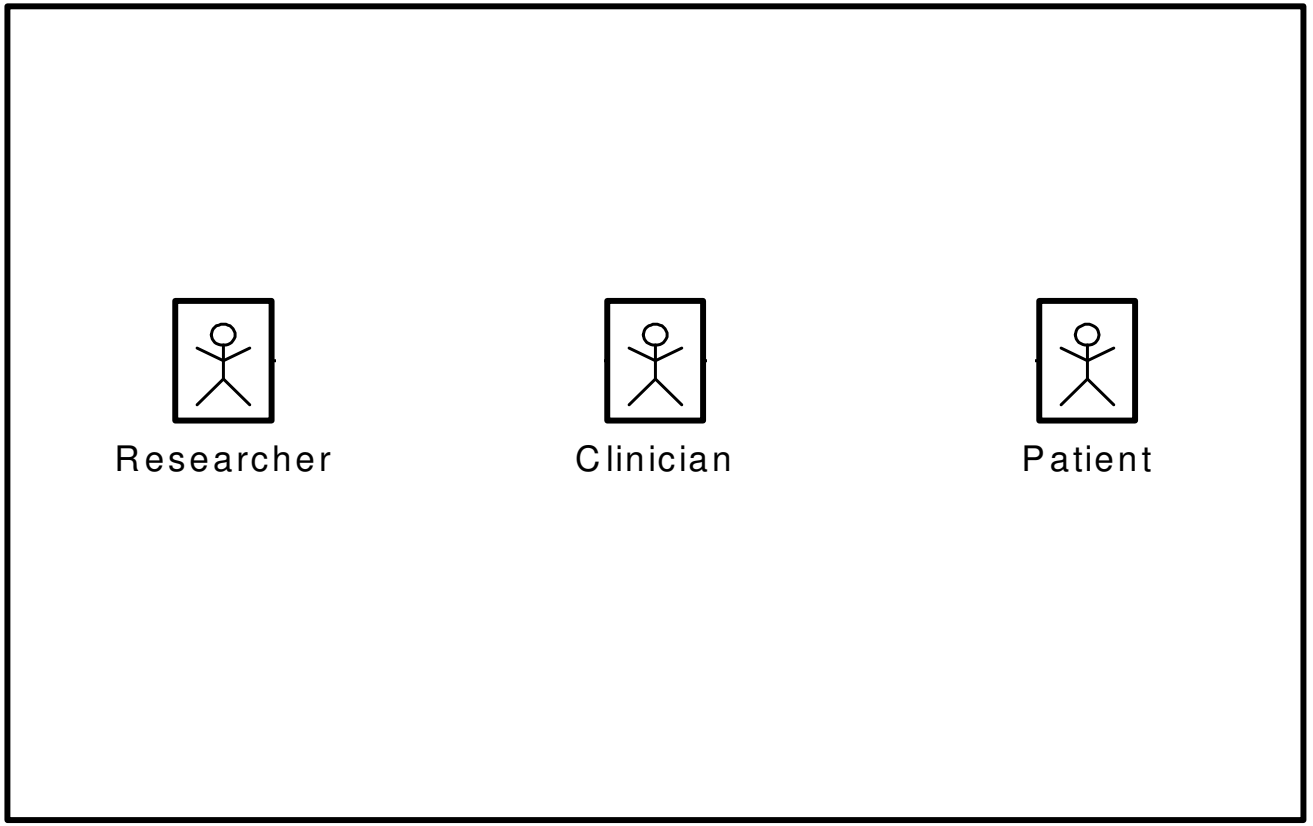
Hype of Hope?

- The volume, velocity, and variety of health data is exponentially increasing
- Health data is scattered in silos, with limited benefits for individuals and society
- Patients want to control their healthcare data and understand them
- Thousands of health applications are targeting specific diseases whilst lacking a holistic view on the patient
- Advances in hardware and software (e.g., machine learning, in-memory databases) enables data processing at large scale

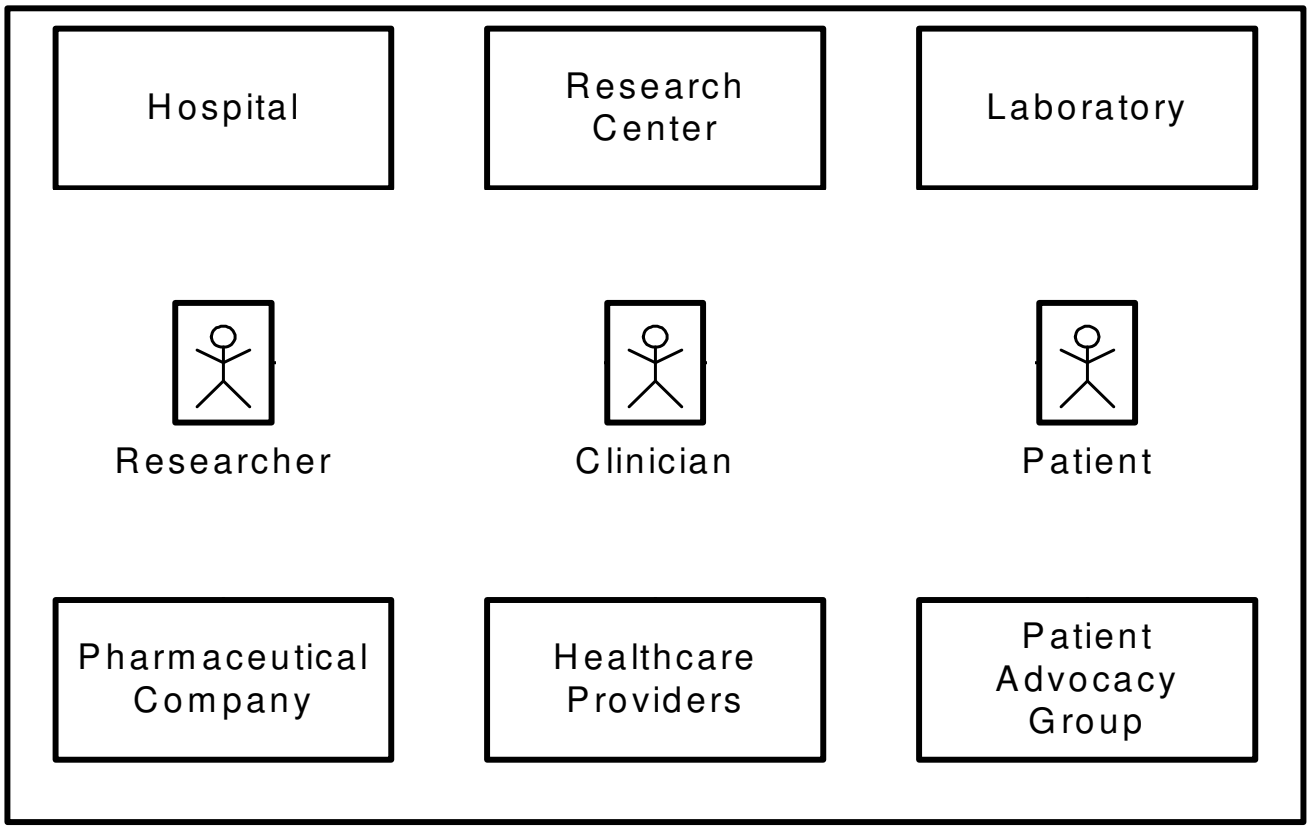
Numbers You Should Know
Comparison of Costs



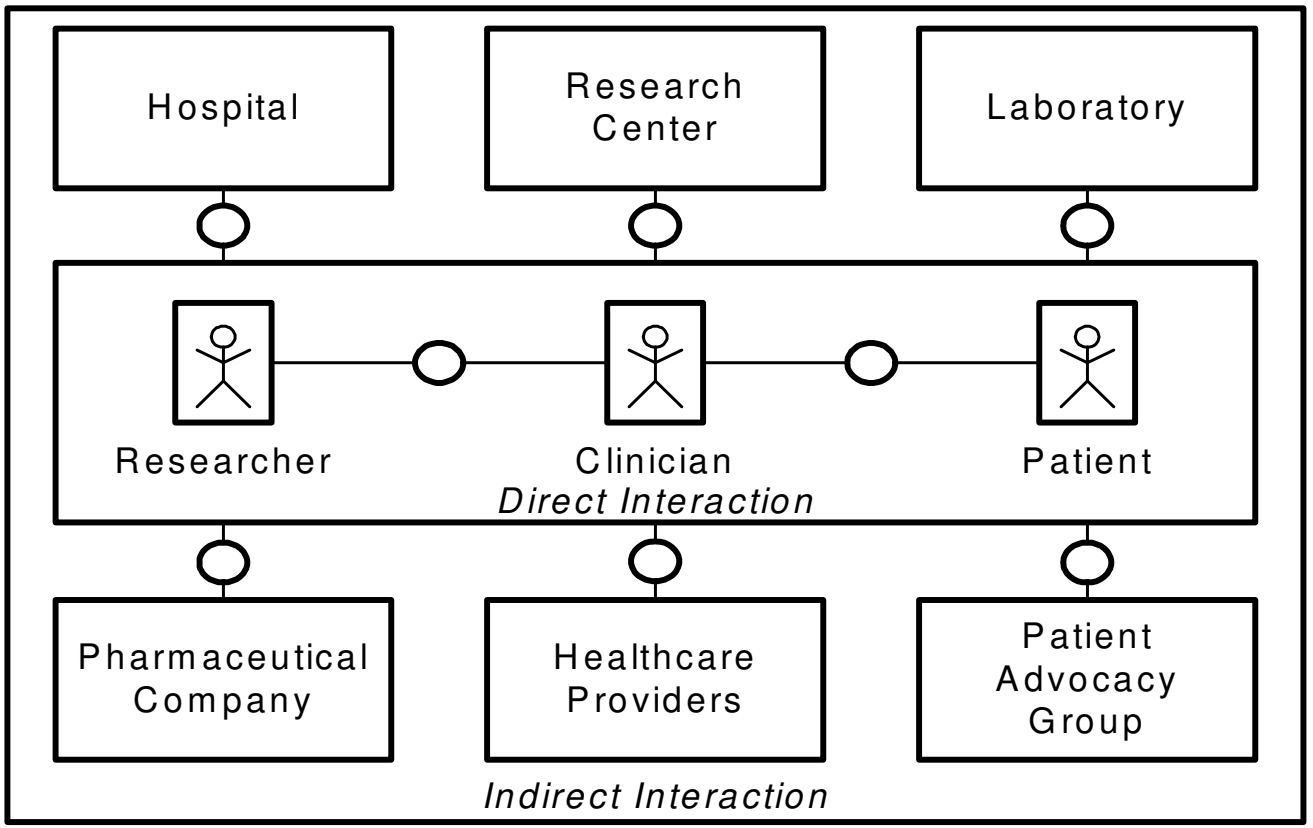
Intelligent Healthcare Networks in the 21st Century?



Intelligent Healthcare Networks in the 21st Century?



Intelligent Healthcare Networks in the 21st Century!



The Setting

Hasso Plattner Institut

- **Patients**

  - Individual anamnesis, family history, and background
  - Require fast access to individualized therapy

- **Clinicians**

  - Identify root and extent of disease using laboratory tests
  - Evaluate therapy alternatives, adapt existing therapy

- **Researchers**

  - Conduct laboratory work, e.g. analyze patient samples
  - Create new research findings and come-up with treatment alternatives

The Challenge
Distributed Heterogeneous Data Sources

Hasso Plattner Institut

Human genome/biological data
600GB per full genome
15PB+ in databases of leading institutes

Human proteome
160M data points (2.4GB) per sample
>3TB raw proteome data in ProteomicsDB

Hospital information systems
Often more than 50GB

PubMed database
>23M articles

Medical sensor data
Scan of a single organ in 1s creates 10GB of raw data

Cancer patient records
>160k records at NCT

Prescription data
1.5B records from 10,000 doctors and 10M Patients (100 GB)

Clinical trials
Currently more than 30k recruiting on ClinicalTrials.gov

Our Vision
Interdisciplinary Tumor Board

Hasso Plattner Institut

Our Goal
Informed Decision Making

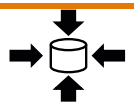
- Can we enable doctors to:
  - Select best treatment options for their patients,
  - Analyze latest diagnostic data about patient's status, and
  - Exchange knowledge with patients to improve quality of living.

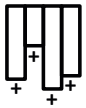
Our Technology
In-Memory Database Technology

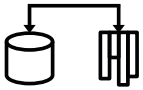

Combined column and row store


Map/Reduce

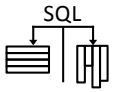

Single and multi-tenancy


Lightweight compression

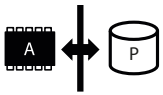

Insert only for time travel


Real-time replication

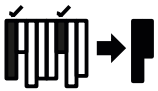

Working on integers

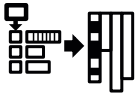


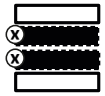
SQL interface on columns and rows



Active/passive data store


Minimal projections


Group key

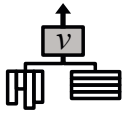

Reduction of software layers


Dynamic multi-threading

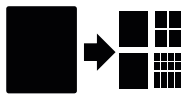

Bulk load of data

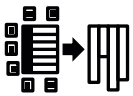

Object-relational mapping

Text retrieval and extraction engine

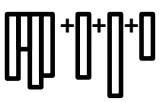


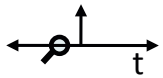
No aggregate tables


Data partitioning


Any attribute as index

No disk

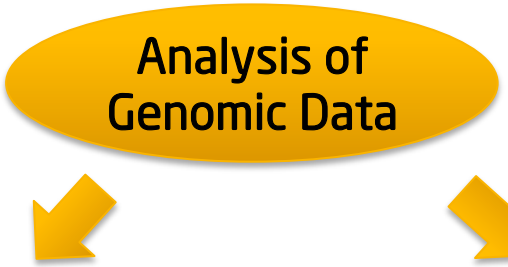

On-the-fly extensibility



Analytics on historical data

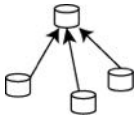

Multi-core/ parallelization

In-Memory Database Technology
Use Case: Analysis of Genomic Data

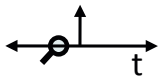


	Alignment and Variant Calling	Analysis of Annotations in World-wide DBs
Bound To	CPU Performance	Memory Capacity
Duration	Hours – Days	Weeks
HPI	Minutes	Real-time
In-Memory Technology	Multi-Core 	Partitioning & Compression 

From University to Market
Oncolyzer



Unified access to formerly disjoint oncological data sources



Flexible analysis on patient's longitudinal data

- Research initiative for exchanging relevant tumor data to improve personalized treatment
- Real-time analysis of tumor data in seconds instead of hours
- Information available at your fingertips: In-memory technology on mobile devices, e.g. iPad
- Interdisciplinary cooperation between clinicians, clinical researchers, and software engineers
- **Honored with the 2012 Innovation Award of the German Capitol Region**

From University to Market
Oncolyzer: Patient Details Screen

- Combines patient's longitudinal time series data with individual analysis results
- Real-time analysis across hospital-wide data using always latest data when details screen is accessed
- https://we.analyzegenomes.com/apps/oncolyzer-mobile-app/

From University to Market
Oncolyzer: Patient Analysis Screen

- Allows real-time analysis on complete patient cohort
- Supports identification of clinical trial participants based on their individual anamnesis
- Flexible filters and various chart types allow graphical exploration of data on mobile devices

From University to Market
SAP EMR: Patient Overview Screen

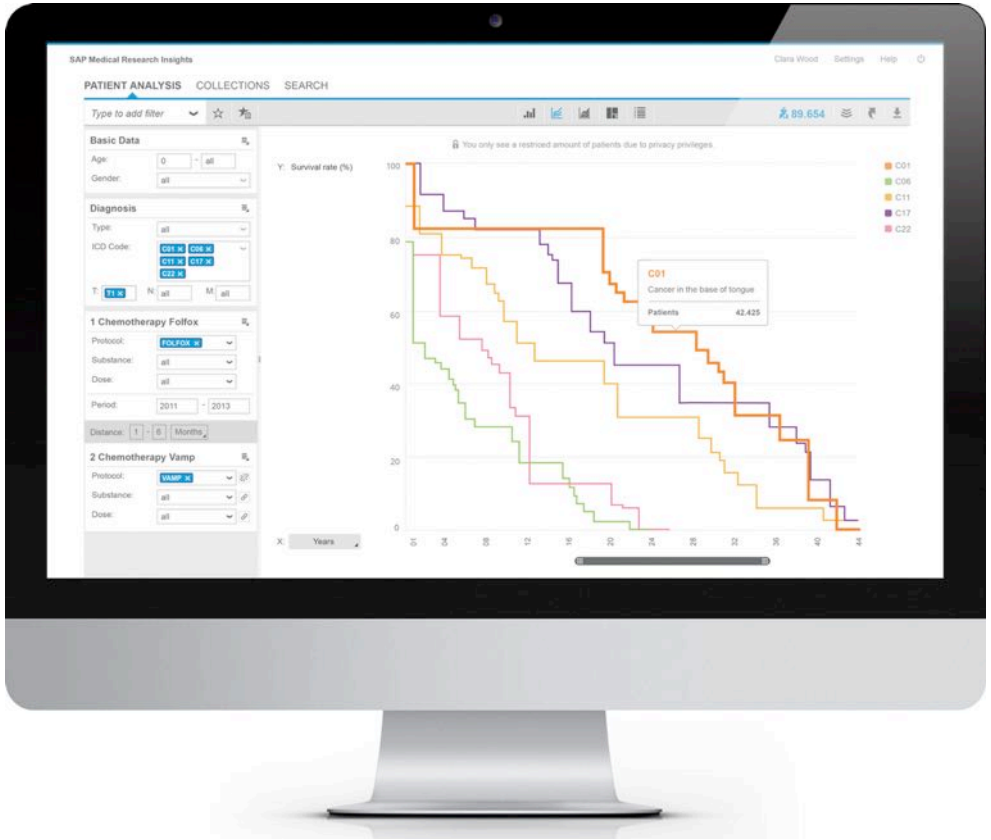
- Shows all patients the logged-in clinician is assigned for
- Provides overview about most recent results and treatments for each patient
- http://global.sap.com/germany/solutions/technology/enterprise-mobility/healthcare-apps/mobile-patient-record-app.epx

From University to Market
SAP EMR: Patient Detail Screen

- Displays time series data, e.g. temperature or BMI
- Allows graphical exploration of time series data

From University to Market
SAP Medical Research Insights

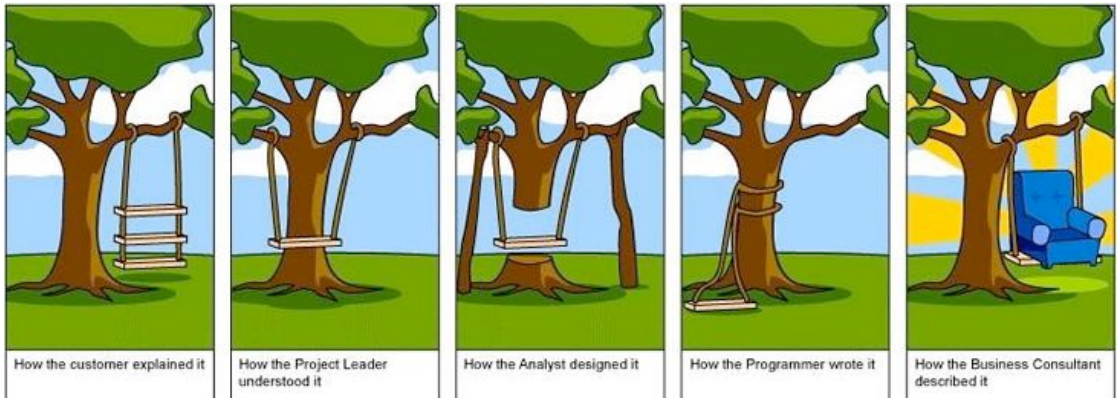
- Flexible combination of medical data
- Enables interactive and graphical exploration
- Easy to use even without specific IT background



Software Requirements in Life Sciences

Hasso Plattner Institut

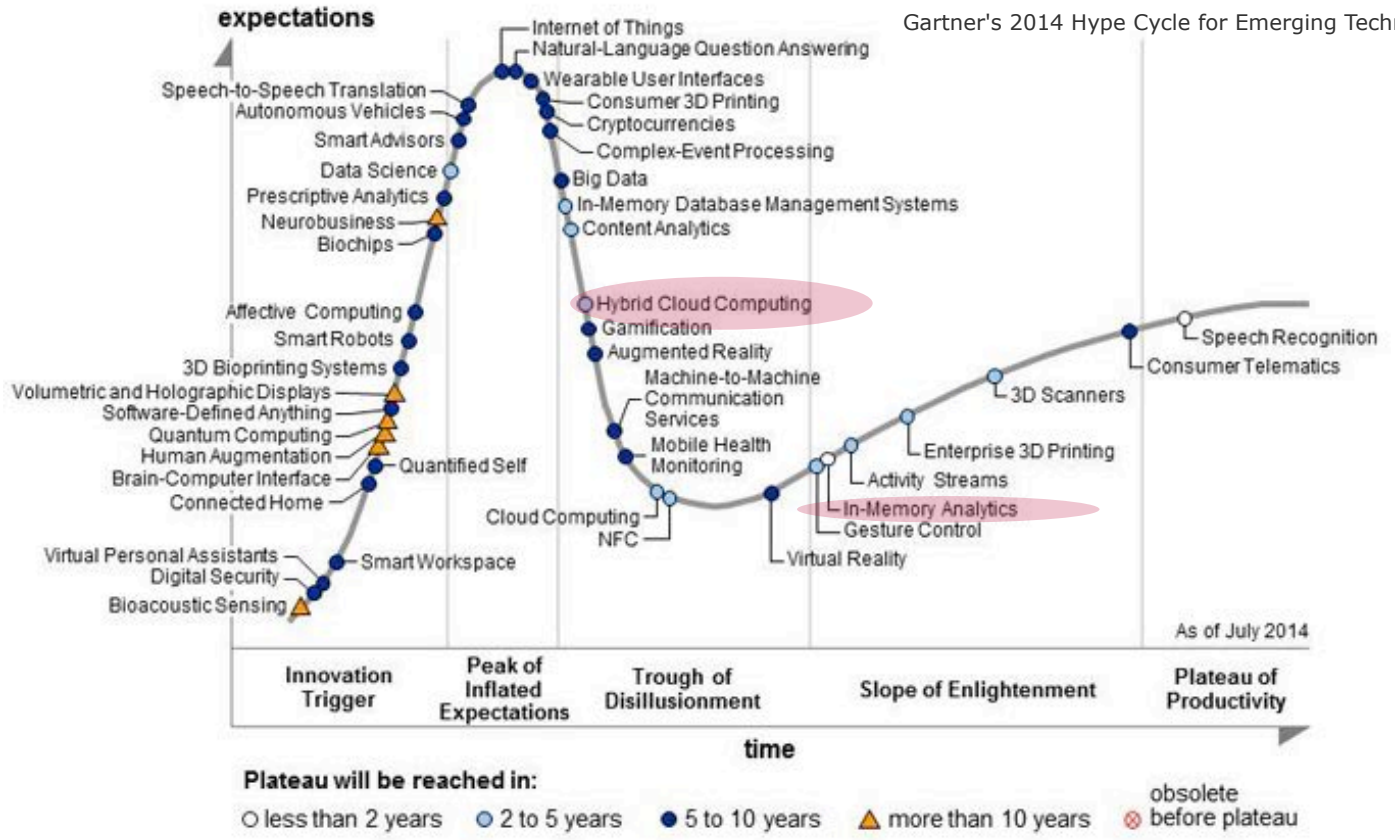
- Requirements
  - Managed services
  - Reproducibility
  - Real-time data analysis
- Restrictions
  - Data privacy
  - Data locality
  - Volume of big medical data



http://stevedempsen.blogspot.de/2013/08/agile-software-requirements-comic.html

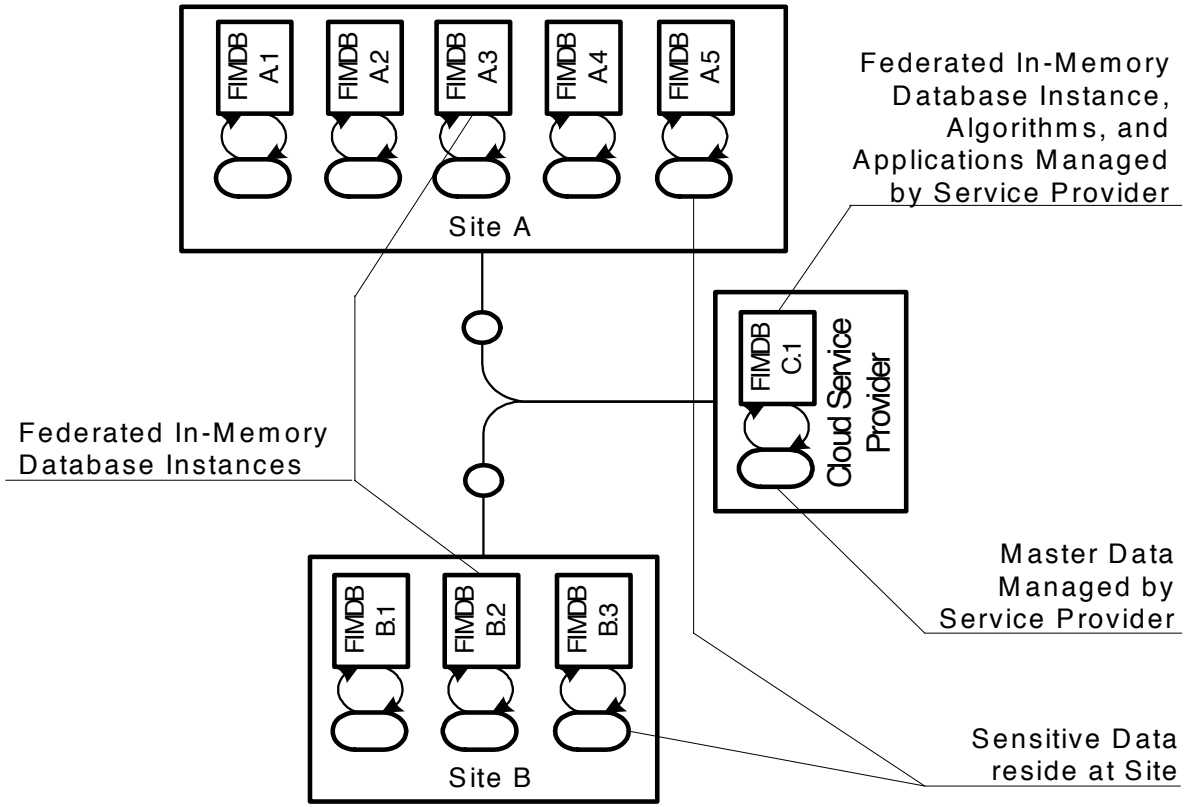
Where are all those Clouds go to?

Hasso Plattner Institut

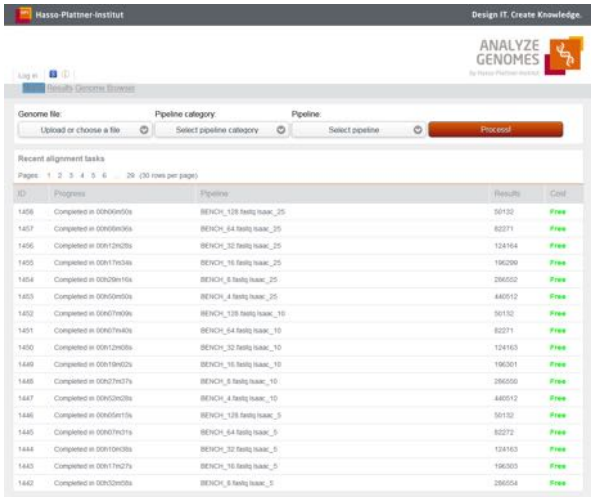


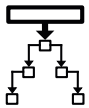
Gartner's 2014 Hype Cycle for Emerging Technologies

Federated In-Memory Database (FIMDB) Incorporating Local Compute Resources



App Example:
From Raw DNA to Variants

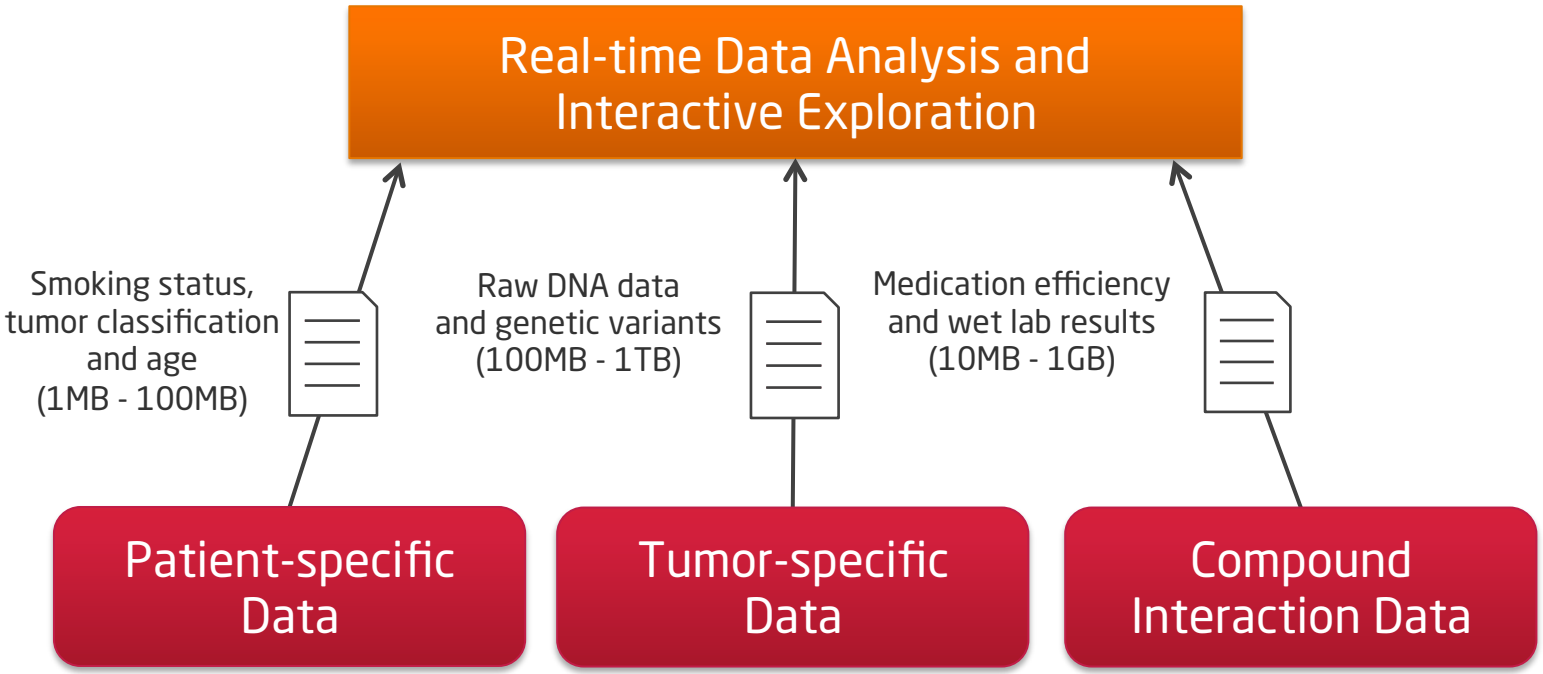




Standardized Modeling and runtime environment for analysis pipelines

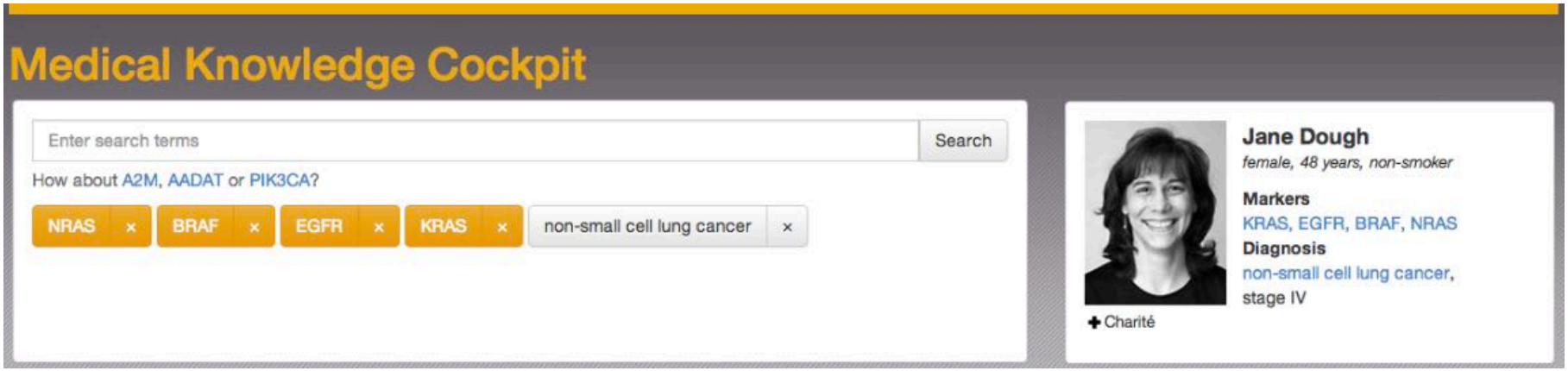
- Control center for processing of raw DNA data, such as FASTQ, SAM, and VCF
- Personal user profile guarantees privacy of uploaded and processed data
- Supports reproducible research process by storing all relevant process parameters
- Implements prioritized data processing and fair use, e.g. per department or per institute
- Supports additional service, such as data annotations, billing, and sharing for all Analyze Genomes services
- **Honored by the 2014 European Life Science Award**

App Example:
Identification of Optimal Chemotherapy



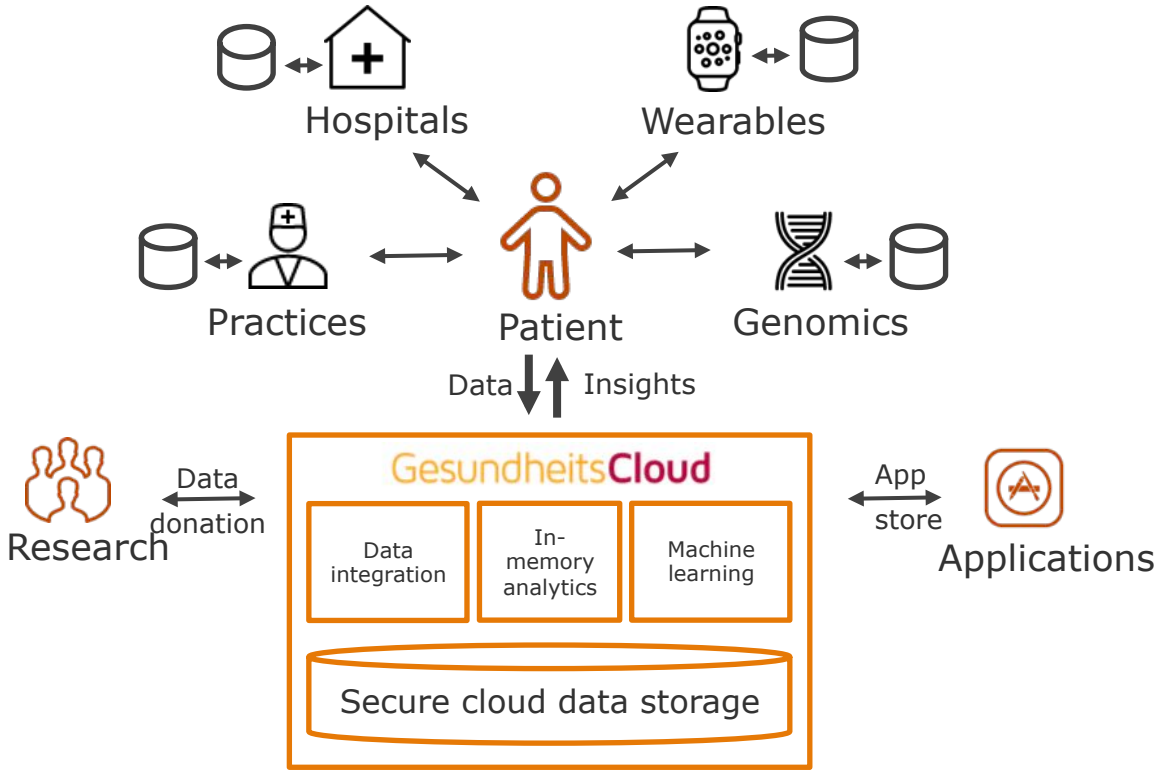
- **Honored by the 2015 PerMediCon Award**

App Example:
Latest Medical Knowledge for Patients and Clinicians



- Query-oriented search interface
- Seamless integration of patient specifics, e.g. from EMR
- Parallel search in international knowledge bases, e.g. for biomarkers, literature, cellular pathway, and clinical trials

App Example: GesundheitsCloud
Combining Distributed Health Care Data Sources



App Example: GesundheitsCloud
Combining Distributed Health Care Data Sources

GesundheitsCloud

Hasso Plattner Institut

Patients

- Improved individualized healthcare provisioning
- Access to personal healthcare data
- Education through domain-specific expert apps
- Data donation

Medical Doctors

- Holistic view on patient anamnesis
- Advanced support for decision making, e.g. machine learning and real-time data analysis
- Improved clinical trial participation

Researchers

- Access to large real-world cohort data for research
- Exploration of society-wide effects

App providers

- Access to secure data processing, analysis, and storage infrastructure
- Use of donated data for improvement and testing of apps
- Access to specific user groups per app

Society: Improved healthcare at lower costs

Do Not Forget to Enroll for the Lecture!

We want you!

Contacts

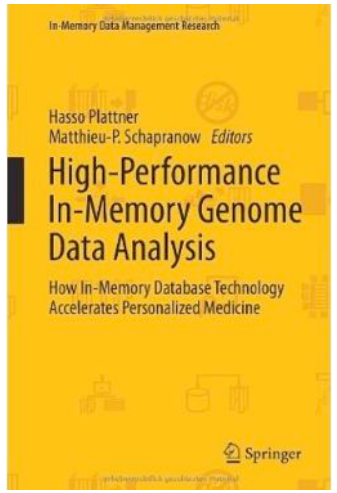
- Milena Kraus
- Harry Freitas da Cruz
- Dr.-Ing. Matthieu-P. Schapranow
- Dr. Matthias Uflacker

✉ <Name>.<Surname>@hpi.de

Enterprise Platform and Integration Concepts (EPIC)
Hasso Plattner Institute
August-Bebel-Str. 88
14482 Potsdam, Germany

https://we.analyzegenomes.com/



Introduction to Digital Health

Data Management for Digital Health, Summer 2017

42