



1st Fall School on Efficient Architectures for Data Science (EADS 2017)

Hasso Plattner Institute
Potsdam | September 18 - 22, 2017

Design IT. Create Knowledge.

Agenda

Data science applications require an adept handling of underlying hardware and software resources to achieve the required scalability and efficiency. To unleash the full potential of data science technologies, the developers' understanding needs to go beyond the mere usage of third-party library interfaces. Emerging data science applications can only benefit from improved hardware acceleration, if compute and storage resources are managed efficiently. Non-uniform memory hierarchies and heterogeneous computing architectures offer enormous opportunities but also pose new challenges for deep learning and other machine learning approaches. The **Fall School on Efficient Architectures for Data Science** will bring together students, researchers and industry practitioners to explore these opportunities and challenges.

The event focuses on hands-on experience and mutual exchange. It features two parts:

- There will be **expert-taught classes** on data science technologies, efficient resource utilization, and advanced optimization approaches.
- The majority of the time will be allocated to a **practical software development and optimization contest**, which will challenge participants to apply their knowledge in a real-world data science scenario.

We are looking forward to an exciting week of inspiration, learning, and hacking.

Monday, September 18, 2017

Room: **HS 2** (Talks), **H-E.11/13** (Hands-on)

9:00 - 9:30	Registration
9:30 - 9:45	Opening of the Fall School on Efficient Architectures for Data Science Prof. Dr. Andreas Polze, Hasso Plattner Institute, Potsdam
09:45 - 10:30	Keynote Prof. Dr. Felix Naumann, Hasso Plattner Institute, Potsdam Big Data Science
10:30 - 10:45	Coffee Break
10:45 - 12:30	Current Applied Research Christian Bartz, Hasso Plattner Institute, Potsdam End-to-End Detection and Recognition Systems for Computer Vision Tim Repke, Hasso Plattner Institute, Potsdam Analysing Business Communication
12:30 - 14:00	Lunch Break

- 14:00 - 14:45 **EADS Challenge**
 Bernhard Rabe, HPI Future SOC Lab, Potsdam
[HPI Future SOC Lab Insights](#)
 Frank Feinbube, Chairman Projects, SAP Innovation Center, Potsdam
[EADS 2017 Challenge Details](#)
- 14:45 - 17:00 **Hands-on: Challenge, Frameworks, Tools**
 Sven Köhler, Lena Feinbube, Carsten Walther, Max Plauth
[Getting started: Infrastructure, Framework, MPI, OpenMP, CUDA, SIMD](#)
- 17:00 - 17:15 **Wrap-up & Feedback**
- 18:00 **Social Event**
[Barbecue at HPI](#)

Tuesday, September 19, 2017

Room: [HS 2 \(Talks\)](#), [H-E.11/13 \(Hands-on\)](#)

- 9:00 - 11:15 **Hands-on: Resource Management and Performance Optimization**
 Thomas Jakobs, Chemnitz University of Technology, Chemnitz
[Optimizations for the Future: Energy and Power Efficiency](#)
 Sven Köhler, Hasso Plattner Institute, Potsdam
[Parallel Programming with MPI](#)
 Felix Eberhardt, Hasso Plattner Institute, Potsdam
[Optimization Guidelines for NUMA Architectures](#)
- 11:15 - 11:30 **Coffee Break**
- 11:30 - 12:30 **Hands-on: Resource Management and Performance Optimization**
 Sven Köhler, Hasso Plattner Institute, Potsdam
[Data-parallel Programming with Vector Instructions](#)
 Max Plauth, Hasso Plattner Institute, Potsdam
[An Introduction to GPU Accelerated Computing](#)
- 12:30 - 14:00 **Lunch Break**
- 14:00 - 17:00 **Hands-on: Challenge Details & Project Time**
 Max Plauth, Felix Eberhardt, Frank Feinbube, Sven Köhler, Carsten Walther
- 17:00 - 17:15 **Wrap-up & Feedback**

Wednesday, September 20, 2017

Room: [HS 2 \(Talks\)](#), [H-E.11/13 \(Hands-on\)](#)

- 9:00 - 10:00 **Fast Abstracts (EADS2017 Attendees)**
 Laura Morgenstern
[Towards a NUMA-aware Task-based Fast Multipole Method](#)
 Kate Hofmann
[A Framework for Evaluation of Computational Cognition](#)
 Cristian Camilo Castellanos
[BDA ArchOps: An Executable Architectural Model for Big Data Analytics](#)
 Sedigheh Eslami
[Privacy for Data Science: The Trade-off Between Privacy and Utility](#)

- 10:00 - 10:45 **Stories from the Industry**
 Dr. Rainer Kerth, Chief Architect, SAP Leonardo ML foundation
[SAP Leonardo ML Foundation: Overview of SAP's Platform for Machine Learning](#)
- 10:45 - 11:00 **Coffee Break**
- 11:00 - 12:30 **Stories from the Industry**
 Megha Agarwal, Data Scientist II, Pivotal Software, Inc
[Agile Data Science](#)
 Dr. Fritz Schinkel, Head of Competence Center Big Data, Product Business, Fujitsu
[Machine Learning for Big Data Analysis](#)
- 12:30 - 14:00 **Lunch Break**
- 14:00 - 17:00 **Hands-on: Project Time**
 Max Plauth, Felix Eberhardt, Frank Feinbube, Sven Köhler, Carsten Walther
- 17:00 - 17:15 **Wrap-up & Feedback**
- 17:20 **Social Event**
[Barbecue at Innovation Center Potsdam](#)

Thursday, September 21, 2017

Room: [HS 2](#) (Talks), [H-E.11/13](#) (Hands-on)

- 09:00 - 12:30 **Hands-on: Project Time**
 Max Plauth, Felix Eberhardt, Frank Feinbube, Sven Köhler, Carsten Walther
- 12:30 - 14:00 **Lunch Break**
- 14:00 - 17:00 **Hands-on: Project Time**
 Max Plauth, Felix Eberhardt, Frank Feinbube, Sven Köhler, Carsten Walther
- 17:00 - 17:15 **Wrap-up & Feedback**
- 17:15 - 21:45 **Project Time**
- 22:00 **Deadline Contest Submission**

Friday, September 22, 2017

Room: [HS 2](#) (Talks), [H-E.11/13](#) (Hands-on)

- 9:00 - 10:00 **Presentation Preparation**
 EADS2017 Attendees
- 10:00 - 12:00 **Project Presentations**
 EADS2017 Attendees
- 12:00 - 13:00 **Lunch & Jury Meeting**
- 13:00 - 13:30 **Reception**
- 13:30 - 14:00 **Award Ceremony & Closing Remarks**
 Prof. Dr. Andreas Polze, Hasso Plattner Institute

Hasso Plattner Institute for Digital Engineering

The Hasso Plattner Institute for Digital Engineering in Potsdam is Germany's university excellence center for IT-Systems Engineering. HPI is the only university institution in Germany to offer the Bachelor's and Master's degree in "IT-Systems Engineering" – a practical and engineering-oriented alternative to a conventional computer science study program. Current enrollment is at approximately 500 students. It carries out research noted for its high standard of excellence in its twelve topic areas. Research work is also conducted at the HPI Research School for PhD candidates as well as at its branches in Cape Town, Haifa and Nanjing.

The HPI School of Design Thinking is Europe's first innovation school for university students. It is based on the Stanford [d.school](#) model and offers 240 places annually for a supplementary study program. Since 2012 the Hasso Plattner Institute has offered Massive Open Online Courses (MOOCs) at its own interactive platform, [openHPI](#). The courses on IT topics have so far reached more than 140,000 users from over 180 countries. Free via the Internet, learners can access didactically prepared multi-media course materials and learn in close exchange with other course participants through social media.



The HPI Future SoC Lab

The HPI Future SOC (Service-Oriented Computing) Lab is a cooperation of the Hasso Plattner Institute and the industrial partners Dell EMC, Fujitsu, SAP and Hewlett Packard Enterprise. Its mission is to enable and promote exchange and interaction between the research community and the industrial partners.

The Lab provides researchers with free of charge access to a complete infrastructure of state of the art hard- and software. This infrastructure includes components, which might be too expensive for an ordinary research environment. The offerings address researchers particularly from but not limited to the areas of computer science and business information systems. Main areas of research include cloud computing, parallelization, and In-Memory technologies.

Future SOC Lab Day – Fall 2017

On **Wednesday, November 15, 2017**, the projects of the previous Future SOC Lab period get a chance to present the results of their research activities. Additionally, selected requesters of new projects can expand their ideas. You are welcome to hand in a project proposal and to apply for the usage of the lab's resources. More information: <https://hpi.de/future-soc-lab>

PROJECT PARTNERS

The image displays a grid of logos for project partners. The logos are arranged in approximately 10 rows and 10 columns. Some of the recognizable logos include Fraunhofer FOKUS, TU Muenchen, Universität Passau, Georg-August-Universität Göttingen, IIMEMAU University of Technology, Technische Universität Braunschweig, Hochschule Pforzheim, Universität des Saarlandes, MIT, Masdar Institute of Technology, Insight, Freie Universität Berlin, Universität Koblenz-Landau, Technische Universität Dresden, Hochschule für Technik Stuttgart, Universität Potsdam, UCL, Universität Leipzig, Universität Mannheim, Universität Würzburg, Fachhochschule Dortmund, Universität St. Gallen, HPI Hasso Plattner Institut, UPC, MPMIC, UNISA, and Universität Passau. The logos are of various shapes and colors, representing different academic and research institutions.

HPI Hasso Plattner Institut
IT Systems Engineering | Universität Potsdam

The image shows the logo for HPI Hasso Plattner Institut. The logo consists of the letters 'HPI' in white on a red square background, followed by the text 'Hasso Plattner Institut' in black. Below this, it says 'IT Systems Engineering | Universität Potsdam'. The logo is centered on a background of a white network graph with grey nodes and lines.

INDUSTRIAL PARTNERS

The image displays the logos for industrial partners. The logos are arranged in a single row. From left to right, they are: SAP (blue and white), Fujitsu (red and white), Dell EMC (blue and white), and Hewlett Packard Enterprise (green and white).

Speakers



Prof. Dr.
Andreas Polze

Hasso Plattner Institute,
Potsdam

Prof. Dr. Andreas Polze is the Operating Systems and Middleware Professor at the Hasso Plattner Institute at University Potsdam, Germany. He is also the speaker of the HPI Research School and member of the steering committee of HPI's Future SOC Lab. Andreas received a doctoral degree from Freie University Berlin, Germany, in 1994 and a habilitation degree from Humboldt University Berlin in 2001, both in Computer Science. At HPI, his research focuses on architectures of operating systems, on component-based middleware, as well as on predictable distributed and cloud computing. Andreas Polze was visiting scientist with the Dynamic Systems Unit at Software Engineering Institute, at Carnegie Mellon University, Pittsburgh, USA, where he worked on real-time computing on standard middleware (CORBA) and with the Real-Time Systems Laboratory at University of Illinois, Urbana-Champaign. Current research interests include Predictable Service Computing, Adaptive System Configuration, and End-to-End Service Availability for standard middleware platforms. Together with Charité, GETEMED, and Deutsche Telekom, he has run the Fontane telemedicine project. Joint research with SAP has investigated porting HANA to new processor architectures.

Big Data Science

While data science describes a broad set of technologies, use cases, systems, and skills, this talk focuses on its data engineering aspects. After an introduction into the general problems and challenges of data engineering in data science undertakings, we focus on the data preparation task, which takes up more than 60% of the time and effort spent on typical data science projects. Therein, we will discuss the problems of data profiling, i.e., the initial extraction of metadata, and those of data quality and their corresponding cleaning methods.



Prof. Dr. **Felix Naumann**

Hasso Plattner Institute,
Potsdam

Felix Naumann studied mathematics, economy, and computer sciences at the University of Technology in Berlin. He completed his PhD thesis on “Quality-driven Query Answering” in 2000. In 2001 and 2002 he worked at the IBM Almaden Research Center on topics of data integration. From 2003–2006 he was assistant professor for information integration at the Humboldt-University of Berlin. Since then he holds the chair for information systems at the Hasso Plattner Institute at the University of Potsdam in Germany. He is editor-in-chief of Information Systems and his research interests are in data profiling, data cleansing, and text mining.



Christian Bartz

Hasso Plattner Institute,
Potsdam

End-to-End Detection and Recognition Systems for Computer Vision

With the breakthrough of the AlexNet at the ImageNet-Competition in 2012, where Krizhevsky et al. showed that a computer vision system based on deep learning can outperform traditional feature engineering approaches by a large margin, deep learning became the driving factor behind state-of-the-art computer vision research. Thanks to the availability of large datasets and hardware capable of performing massive parallel computations, the usage of deep learning in computer vision is steadily growing.

In this talk I will provide an overview of current state-of-the-art end-to-end recognition systems for both object detection/recognition and scene text detection/recognition. We will have a look at the problems arising from the lack of labelled data for the supervised training of deep models, and the need for an efficient usage of the hardware for the faster generation of experimental results, and also shortening run-time, especially on mobile devices.

Christian Bartz studied at the Hasso Plattner Institute in Potsdam and the University of Technology Sydney (Australia) and received his master degree in IT-Systems Engineering from the Hasso Plattner Institute in 2016. Since then working as a PhD student with the Multimedia Analysis and Deep Learning Team at the chair for Internet-Technologies and Systems at the Hasso Plattner Institute. In his work he is concentrating on solving computer vision problems with the help of deep learning, especially in the domain of end-to-end scene text detection and recognition.

Analysing Business Communication

Internal documents report almost all discussions and decisions of a business, as well as their cooperation with other companies through contracts, receipts, or protocols. Establishing an overview of such a corpus is almost impossible due to the sheer quantity and diversity of all the unstructured documents. After major leaks such as the Panama-Papers, Journalists want to quickly gain insights by identifying salient entities and subjects as well as their relationship over time. Special investigators also face these challenges during internal audits or legal prosecution.

This talk outlines some of the challenges we face in our research project, where we develop a system, which helps to explore such corpora by automatically extracting core information and structuring it to create interactive visualisations. We work closely together with our partner from the financial sector, where this system has to prove its flexibility in practice: Underlying deep learning models need to generalise, often even involving scanned documents of different quality. It also has to be able to efficiently process varying amounts of data at scale.



Tim Repke

Hasso Plattner Institute,
Potsdam

As a PhD student at the Hasso-Plattner-Institute, Potsdam, Germany, Tim Repke develops methods that help gain insights from large and diverse collections of documents and make them explorable. This work touches the fields of text mining, document classification, named entity extraction and linking, as well as social network graph analysis. He found his interest in information extraction and unstructured data working in a London-based startup, where he built a general web scraper, which became the topic of his master's thesis on extraction of citation data from arbitrary webpages based on visual cues.



Bernhard Rabe

HPI Future SOC Lab, Potsdam

HPI Future SOC Lab Insights

The HPI Future SOC Lab is a cooperation of the HPI and the industrial partners Dell EMC, Fujitsu, SAP and Hewlett Packard Enterprise founded in 2010. The Lab offers researchers free of charge access to a complete industrial infrastructure landscape. These might be too expensive for an ordinary research environment. The offerings address researchers particularly from but not limited to the areas of computer science and business information systems. Main areas of research include cloud computing, parallelization, and In-Memory technologies.

The talk addresses available resources, usage processes and technical background for effective EADS attendance.

Bernhard has studied computer science at Humboldt University Berlin and received his diploma in 2002. 2001 he was part founding base of the Operating Systems and Middleware Group at the Hasso Plattner Institute and leads the technical matters of the Future SOC Lab since its foundation in 2010.

EADS 2017 Challenge Details

This talk will introduce you to the challenge we prepared for your week at the EADS 2017. In addition to your task, we will discuss the grading metrics, programming template and the submission process. Furthermore, we will introduce you to some of the tools and frameworks that will help you to reach outstanding performance with your machine learning solution.



Frank Feinbube

Chairman Projects, SAP
Innovation Center, Potsdam

After studying IT Systems Engineering at the HPI, Frank joined the Operating Systems and Middleware group of Prof. Polze. In his thesis, he worked on the acceleration of operating systems and his research ranges from automatic optimization of task and data placement to Machine Learning for data partitioning. In 2017 Frank joined the Chairman Projects team of the SAP Innovation Center Network.



Thomas Jakobs

Chemnitz University of
Technology, Chemnitz

Optimizations for the Future: Energy and Power Efficiency

In scientific computing and its economic applications the grand goal for programmers is to optimize the performance of a program. This mainly stands for a reduced execution time or the calculation of a larger problem in the same time. Modern computer systems already are limited in their delivered performance by effects arising from physical boundaries with the use of electricity. The dark silicon effect is a famous effect which prevents a higher processor frequency due to heat dissipation. In future computer systems there are and will be more and more tasks to solve to guarantee a certain degree of performance that will deliver classical optimization goals only as secondary goals. The optimizations of applications in any area of modern life will have to consider the effects on the energy and power consumption of the system.

Thomas Jakobs is a research assistant at the chair of Practical Computer Science at the Faculty of Computer Science at Technische Universität Chemnitz. His current research covers energy and power aware vectorization and the usage of hardware transactional memory in scientific and parallel computing environments.

Parallel Programming with MPI

The Message-Passing Interface (MPI) is a standard that allows programmers to write portable parallel applications. These applications can be both run on shared memory workstations as well as distributed across large computing clusters. MPI can serve as glue for your heterogenous computing platforms.

This talk is an introduction to the concepts of shared-nothing parallelism and the MPI-API. You'll be shown how to spawn your programs across multiple computing nodes using the OpenMPI implementation. Parallelisation strategies for typical data science and machine learning problems are discussed.



Sven Köhler

Hasso Plattner Institute,
Potsdam

Sven Köhler is currently working as Ph.D. student with Prof. Andreas Polze at the Hasso Plattner Institute, Potsdam. His master's thesis focussed on accelerators for parallel data processing and cryptography on the IBM POWER8 platform. Further he has a background in engineering complex software systems, computer vision and building interactive devices. In his spare time he mentors a CoderDojo and introduces school kids to the marvels of computing.



Felix Eberhardt

Hasso Plattner Institute,
Potsdam

Optimization Guidelines for NUMA Architectures

Our mental model of multiprocessor computer systems consists of multiple processors connected via a shared bus to one block of memory. The memory access of those systems is uniform in the sense that every core has the same latency and bandwidth to every memory address in the system. However today's server systems are non-uniform memory access systems with multiple processors having their own share of the whole memory and coupled by an arbitrary interconnection network. The memory accesses on those systems have different costs depending on the distances between the accessing thread and the memory location. Bad placement of threads and data can have severe impact on the overall performance of the system due to congestions of the underlying interconnection network. To date best practices and optimization techniques for parallel programming focus either on parallel shared memory systems with the assumption of uniform memory access or distributed systems with a message passing paradigm. But hierarchical NUMA systems have characteristics of both system types. It is not clear what programming model is best suited for the different data access patterns. In this talk I will give a short introduction to the evolution of hierarchical NUMA systems and major problems regarding placement of threads and data. We will have a look into how to observe the runtime behaviour of applications on NUMA systems and discuss performance measurements of selected workloads.

Felix Eberhardt is a researcher at the Operating Systems and Middleware Group at the Hasso Plattner Institute. In 2014 he received his M.Sc. in IT Systems Engineering at the Hasso Plattner Institute. His master's thesis focused on analyzing the runtime behaviour of SAP HANA on NUMA systems. He is currently involved in a project with IBM dealing with analyzing tools and optimization strategies for workloads on hierarchical NUMA machines. In an EU project (SSICLOPS) he is researching the efficient and secure placement of workloads in cloud computing environments.

Data-parallel Programming with Vector Instructions

In-core hardware accelerators are a worthwhile mean to increase compute performance. CPU vendors forge compute-intensive functions into hardware to shorten runtime, increase throughput or reduce energy consumption compared to pure software implementations. This talk shows at the example of the x86_64 platform how Single-Instruction-Multiple-Data (SIMD) instruction sets (i.e. SSE and AVX) can be used to speed up computation on a single processor. Common traps are highlighted and implementation patterns shown.



Sven Köhler

Hasso Plattner Institute,
Potsdam

Sven Köhler is currently working as Ph.D. student with Prof. Andreas Polze at the Hasso Plattner Institute, Potsdam. His master's thesis focussed on accelerators for parallel data processing and cryptography on the IBM POWER8 platform. Further he has a background in engineering complex software systems, computer vision and building interactive devices. In his spare time he mentors a CoderDojo and introduces school kids to the marvels of computing.



An Introduction to GPU Accelerated Computing

GPU compute devices have become very popular for general purpose computations. This talk aims at providing an entry point for developing GPU-accelerated applications using CUDA. For that purpose, the general characteristics of GPU hardware are detailed, and the properties of CUDA are introduced by walking through a simple example.

Max Plauth

Hasso Plattner Institute,
Potsdam

In 2014, Max Plauth obtained his M.Sc. in IT Systems Engineering from the Hasso Plattner Institute at the University of Potsdam in Germany. In his master's thesis, he studied the benefits of leveraging graphics processors for computational tasks in audio signal processing scenarios. Subsequent to his graduation, he joined the workforce of the IVU Traffic Technologies AG as a software engineer. He contributed to a software product that supports the daily workflows of transport companies by aggregating large data volumes. After having gained one year of industry experience, he returned to the academic world in February 2015 and started his PhD in the Operating Systems and Middleware Group at the Hasso Plattner Institute. His research interests include parallel computing in the context of cloud computing. More specifically, he is turning his focus on exploiting the characteristics of different compute device classes in order to perform tasks more efficiently.

Fast Abstracts

Laura Morgenstern Chemnitz University of Technology, Chemnitz

Towards a NUMA-aware Task-based Fast Multipole Method

Kate Hofmann Hasso Plattner Institute, Potsdam

A Framework for Evaluation of Computational Cognition

Cristian Camilo Castellanos University of Los Andes, Bogotá, Colombia

BDA ArchOps: An Executable Architectural Model for Big Data Analytics

Sedigheh Eslami Hasso Plattner Institute, Potsdam

Privacy for Data Science: The Trade-off Between Privacy and Utility

We also welcome contributions of EADS2017 attendees relevant to data science and related fields – they will have the opportunity to introduce new ideas, work in progress, and application or development experience, discuss their research with other researchers and receive feedback from the audience.

SAP Leonardo ML Foundation: Overview of SAP's Platform for Machine Learning

Dr. Rainer Kerth

Chief Architect, SAP Leonardo
ML foundation

This talk will discuss the approach that SAP has chosen to productize Machine Learning functionality in the cloud. To support heavy compute workloads typical for Deep Learning, the development team decided early on to extend the existing cloud platform at SAP and delegate core ML tasks into a Kubernetes cluster, deployed next to the SAP Cloud Platform. This approach allows the ML platform to support long running processes with large memory footprints, such as ML training jobs, which otherwise are difficult to integrate into the existing software infrastructure at SAP. We will describe the technical challenges that we encountered when executing on this overall approach as well as our solutions to some of these challenges.

Dr. Rainer Kerth is the Chief Architect for SAP's Machine Learning platform, known as SAP Leonardo ML foundation. Mr Kerth is with SAP since 2008 and previously worked in SAP's Central Architecture team on various platform infrastructure questions. In previous positions at IBM, Mr Kerth was responsible for introducing an Enterprise Service Bus at a large automotive customer, defining a platform for Straight Through Processing in Financial Services as well as leading the RFID product development at IBM. He also helped to define the architecture of the WebSphere Application Server in its early days and contributed to various industry standard organizations, such as JCP and the Auto ID Center. Mr Kerth holds a PhD in Mathematics from the University of Paris 7 as well as a Diploma from the Technical University in Berlin

Agile Data Science

At Pivotal Labs, we are helping our clients build smart apps driving real time actions using Data Science. The talk will focus on how Data Science is done at Pivotal Labs. It will provide an introduction to applying agile and extreme programming practices to data science and how DS models can be put in production in a robust and timely fashion. The talk will also provide a glimpse on how Pivotal used ML to help large multinational companies on their digital transformation journey.



Megha Agarwal

Data Scientist II, Pivotal
Software, Inc

Megha Agarwal is currently working as a Data Scientist at Pivotal Labs, London. As a Data Scientist, she is responsible for helping the clients identify and deliver value from their data. She is focused on developing smart apps by applying machine learning and statistics. She has done her Masters in Machine Learning and HPC from University of Bristol, UK.



Dr. Fritz Schinkel

Head of Competence Center
Big Data, Product Business,
Fujitsu

Machine Learning for Big Data Analysis

The growth in data production results in increased demand for data analytics. To keep track with the growing the traditional style of modelling is complemented and will sometimes be replaced by machine intelligence and automated model learning. E.g. the rate of text production in social media, web sites, mail etc. requires new methods of content reception. Classical methods like key word tagging can be enhanced by deeper text analysis to figure out topics handled in text and make good suggestions for most relevant documents. In the manufacturing industry the diversity of produced components requires diversity of models which can make analysis quite tedious and expensive. Machine learning algorithms like clustering can find the mayor categories which require different modelling, and therefore avoid fine granular models distinguishing cases with common behavior. The talk presents project examples for both applications of machine learning to analyze big data relevant for the industry.

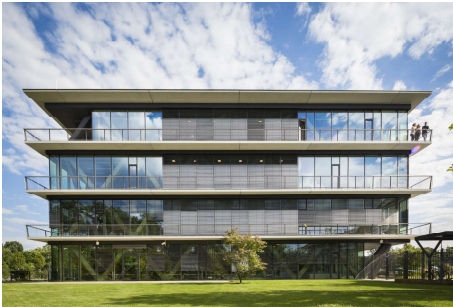
Fritz is heading a big data competence center providing big data solutions and services. His special interests are methods and algorithms for big data analysis. Starting his career in IT industry in 1991 Fritz worked in development, architect and management functions on compilers, CASE tools, web and portal technology, dynamic datacenter management, cloud platform, private cloud and big data program. He has over 25 years of experience in the IT industry working for Siemens, Nixdorf, Fujitsu Siemens and now for Fujitsu. His degree in mathematics Fritz received in 1990 from the Leibniz University in Hannover. In 2015 Fritz has been recognized as one of the Fujitsu Distinguished Engineers.

Social Event - BBQ at Innovation Center Potsdam

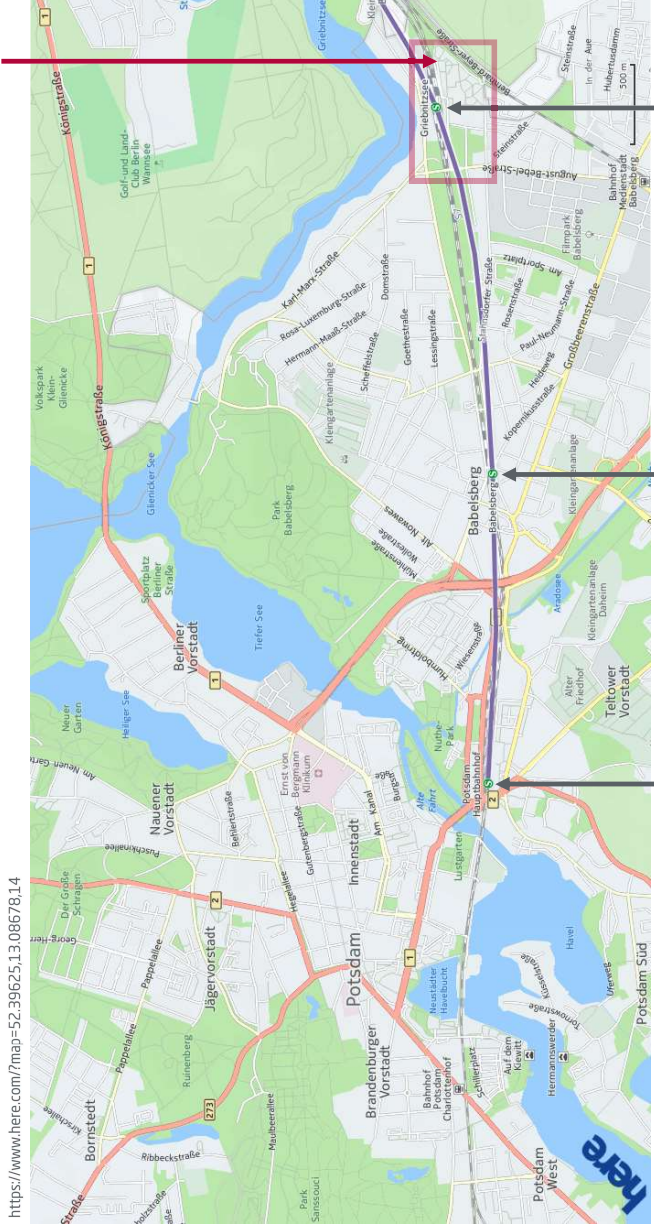
The social event on Wednesday will take place at SAP Innovation Center Potsdam (ICP). Located in the beautiful north of Potsdam next to the Jungfernsee, there will be a guided tour with demos of the latest research and a subsequent barbecue. The shuttle will leave from HPI at 17:30.

As part of SAP's Products & Innovation organization, the SAP Innovation Center Network (ICN) combines the best of two worlds: the creativity and agility of start-ups with the backbone of a world market leader in business software. The ICN creates new growth for SAP by pioneering new markets, disruptive technologies and future business trends. We build innovative solutions in the fields of machine learning, blockchain, and conversational applications and design the future of work with software for the purpose-led enterprise.

The first SAP Innovation Center was established in 2011 in Potsdam. Today, the network connects more than 500 engineers, designers, product experts, and business developers working on a diverse range of projects in 12 locations and nine different countries. Interdisciplinary teams turn creative thinking into innovative solutions with market impact.

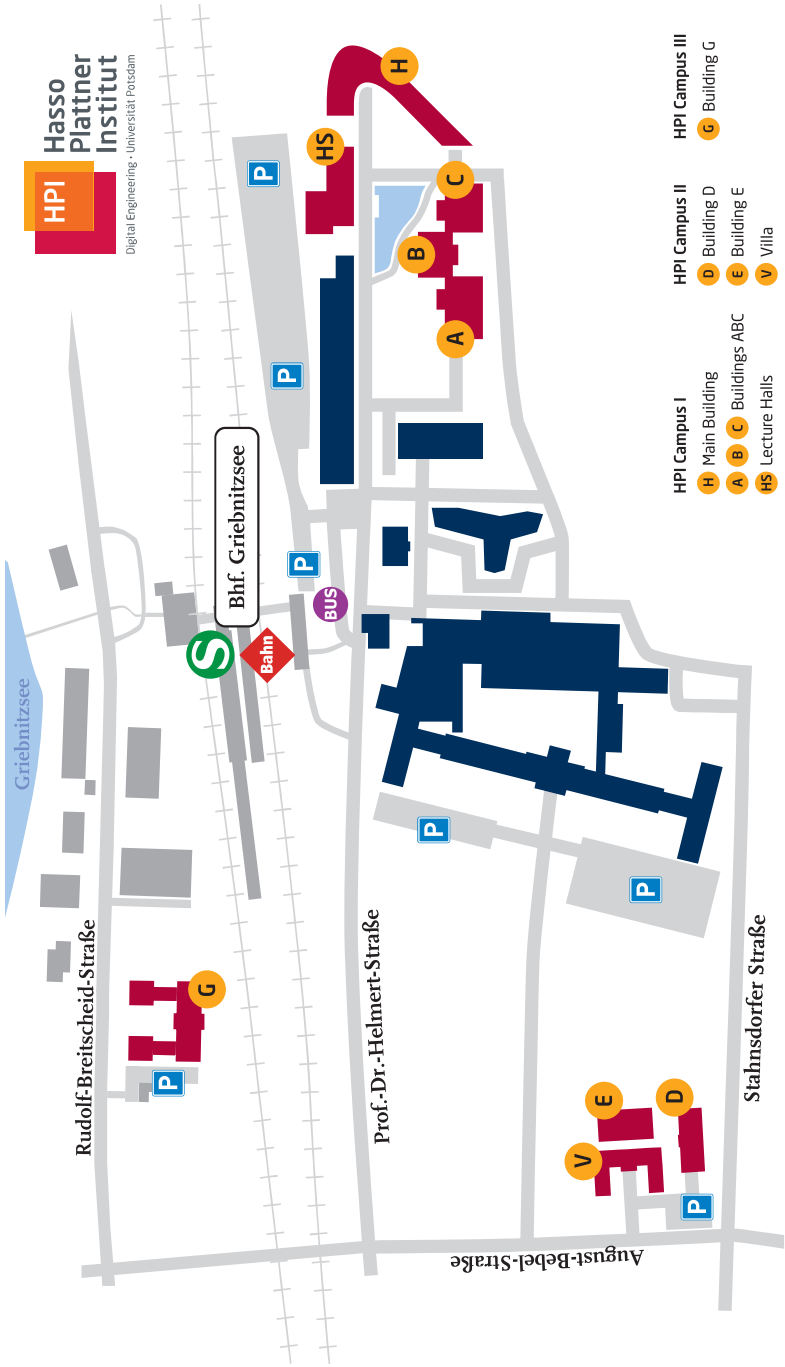


Maps



<https://www.here.com/7map=52.39625,13.08678,14>





HPI
Hasso Plattner Institut
 Digital Engineering · Universität Potsdam

- HPI Campus I**
 - H Main Building
 - A B C Buildings ABC
 - HS Lecture Halls
- HPI Campus II**
 - D Building D
 - E Building E
 - V Villa
- HPI Campus III**
 - G Building G

Notes

Contact

**Hasso Plattner Institute
for Digital Engineering
University of Potsdam**

Campus Griebnitzsee
Prof.-Dr.-Helmert-Straße 2-3
D-14482 Potsdam, Germany

Phone +49 331 5509-0
E-Mail hpi-info@hpi.de
Web <https://www.hpi.de>