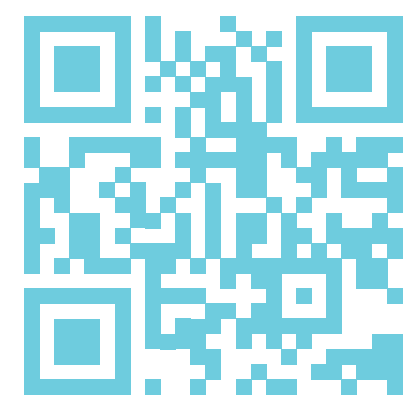


WELCOME TO THE DATA SYSTEMS LAB

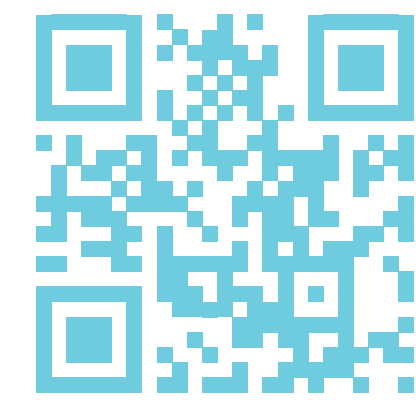
Dear Students,
The Big Data Analytics for Earth Observation (BigEarth) Group, the Data Integration und Data Preparation (D2IP) Group, the Big Data Engineering (DAMS) Group, the Data Engineering for ML (DEEM) Group and the Database Systems and Information Management (DIMA) Group at TU Berlin offer numerous opportunities for you to learn, grow, and develop. Incidentally, BigEarth, D2IP, DAMS, DEEM and DIMA are members of the Data Systems Lab (DASL) in the Berlin Institute for the Foundations of Learning and Data (BIFOLD), a German National Center for Artificial Intelligence. This poster was created to inform you about our educational programs, course offerings, thesis opportunities, and prospective career paths. It is particularly informative for those interested in pursuing a Master's or PhD with a concentration in data management, big data engineering or technologies and systems for data science. Our curriculum is specially designed to ensure sound theoretical knowledge, supplemented with hands-on lab sessions, development projects, and seminars to deepen understanding. Furthermore, our colloquia (Bachelor's/Master's Colloquium and DIMA Research Colloquium) expose you to the current research being undertaken by our scientists and guest speakers from academia and industry. Upon completing our courses, you will possess the foundational, technological, and systems skills you will need to pursue a career in database systems, information management, big data engineering, and technologies and systems for data science. We encourage you to visit and learn more about us.



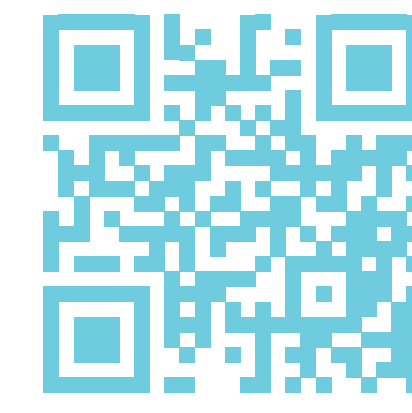
www.tu.berlin/en/d2ip



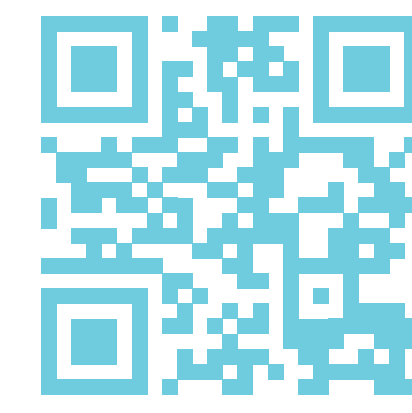
www.tu.berlin/en/dams



www.rsm.berlin

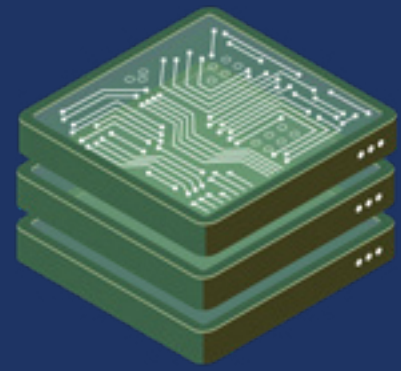


www.tu.berlin/en/dima



www.deem.berlin

DATA SCIENCE & ENGINEERING TRACK



The Data Science and Engineering Track enables students pursuing a Master's Degree in Computer Science, Computer Engineering, or Information Systems Management to specialize and develop expertise in data analytics. In order to fulfill the requirements of this specialization area, students from these aforementioned programs will need to select from an approved set of courses in three core competencies, namely, (1) data analytics, (2) scalable data management, and (3) an application area.

www.tu.berlin/en/dima/analytics/
data-science-and-engineering-track

BACHELOR'S & MASTER'S THESIS OPPORTUNITIES

Students interested in pursuing a thesis in data systems should possess outstanding programming skills in C++, Java, or Scala, deep knowledge in database systems (e.g., IBM DB2, Oracle) or big data analytics systems (e.g., Flink, Spark), basic knowledge in the use of an IDE (e.g., Eclipse, IntelliJ), and basic knowledge in the use of a distributed version control system (e.g., SVN, Git). **Furthermore, to conduct a**

BACHELOR'S THESIS

Students should have successfully completed ISDA and DBPRA (at a minimum) with a grade of good or better and advanced Bachelor's courses offered by the chairs of the Data Systems Lab, in particular, a seminar and a project. **Examples of recent theses include:**

- Investigating efficient co-processing in multi-query stream processing
- Continuous Query Execution Under Continuous Infrastructure Evolution
- Parallelizing Raha and Baran using Dask
- Ordering datasets for efficient or effective error detection
- TPCx-AI Benchmark Implementation on Apache SystemDS
- Adaptive Join Processing as a DuckDB Extension
- Compression of Remote Sensing Images based on Generative Adversarial Networks
- Image Search and Retrieval from Remote Sensing Archives based on Similarity Ranking Functions

MASTER'S THESIS

Students must have successfully completed DBT and DBTLAB (at a minimum) with a grade of good or better and advanced Master's courses offered by the chairs of the Data Systems Lab, in particular, a seminar and a project. **Examples of recent theses include:**

- Towards Unikernel-based Stream Processing
- Towards A GPU-Accelerated Stream Processing Engine Through Query Compilation
- Understanding the Characteristics and Robustness of Data Preparation Pipelines for Machine Learning
- Towards Amnesiac Recommender Systems: Efficient Incremental and Decremental Updates for Next Basket Recommendation
- Self-supervised Data Cleaning in Data Lakes
- Quality-Driven Union Table Search
- Scalable Computation of Shapley Additive Explanations
- Discovering and Generating Rewrites for ML Systems
- Explainable Artificial Intelligence for Multi-Label Remote Sensing Image Classification
- Multi-Query, Cross-Modal and Scalable Content-Based Image Retrieval in Remote Sensing

Moreover, depending on the thesis topic, additional knowledge may be required (e.g., compiler technology, distributed systems, networking, operating systems, systems programming, machine learning).

BACHELOR'S COURSES

Module's Title	Description	ECTS	Term	Type	Level	Language
ISDA Information Systems & Data Analysis	Learn the concepts of information management using (relational) database systems from the perspective of an application developer.	6	SS	IV	Intro	GER
DBPRA Database Practical Hands-on Training	Intensify practical skills in designing, implementing, and administrating relational databases using concrete application examples.	6	WS/SS	PR	Intro	GER/ENG
Practical: Datasystems - Machine Learning Pipelines	Apply software engineering methods to solve a complex practical task and utilize common methods for implementing selected components of machine learning pipelines.	6	WS	PR	ADV	ENG
Programming Project: Data Systems - Data Integration Instance	Learn how to apply software engineering methods to solve a complex data engineering task, and gain experience on working in a team on a collaborative project.	6	WS/SS	PR	ADV	ENG
PPDS Programming Project: Data Systems	Work on a given project in the context of implementing database systems. Learn how to prototype development, the systematic handling of version management, test-driven development, design documentation, and runtime experiments and improvements.	6	WS/SS	PR	ADV	GER
DBPRO Database Project	Develop an information system jointly with a team along a classical development process, including the functional specification, modeling, implementation and demonstration of the system.	6	WS/SS	PJ	ADV	GER
LDE Large-scale Data Engineering	In this combined seminar/project module, you will learn about scientific reading and writing and create prototypes of programming projects in data and ML systems, in the context of big data engineering.	12	WS/SS	PJ + SE	ADV	ENG
DBSEM Seminar on Advanced Topics in Database and Information Systems	Learn about the core elements of a technical presentation, learn how to properly present an advanced scientific topic drawn from the database systems or technologies and systems for big data management and data science literature, and sharpen critical thinking skills.	3	WS/SS	SE	ADV	ENG
BTC Bachelor's Thesis Colloquium in Data Management Systems	Covers a presentation on open thesis topics at DIMA, a presentation and discussion on the expectations and evaluation criteria of Bachelor's Theses, topic selection, structuring and writing a thesis proposal, including problem statement, solution approach, experimental design, evaluation, and implementation plan as well as project and time management over the course of a Bachelor's Thesis.	3	WS/SS	CO	ADV	ENG
BT Bachelor's Thesis	Create a scientific work that solves a data management, big data engineering, or technologies and systems problem: (i) surveying related work, (ii) stating the research problem, (iii) defining the scope, (iv) specifying a solution approach and methodology, (v) differentiating the solution from the state-of-the-art, and (vi) showing the effectiveness and efficiency of your solution.	12	WS/SS	THESIS	ADV	GER/ENG
Data Science for Computer Science	This course introduces you to the fundamental concepts of data science, focusing on data analysis to understand real-world phenomena. Develop skills in data-centric programming, statistical inference, and gain hands-on experience with data collection, preparation, and ethical considerations in data science.	6	SS	VL + TUT	ADV	ENG
DIA Data Integration and Large-Scale Analysis	Learn about major data integration architectures, key techniques for data integration and cleaning, as well as methods for large-scale, i.e., distributed, data storage and analysis.	6	WS	VL + UE	ADV	ENG

MASTER'S COURSES

Module's Title	Description	ECTS	Term	Type	Level	Language
DBT Database Technology	Learn both the fundamentals of data processing in traditional single-node database systems and how to scale out these techniques to huge amounts of data in large-scale, distributed environments.	6	WS	IV	Intro	ENG
DBTLAB Database Technology Lab	Learn how to implement components of a database system. You will create a working SQL query processor that can answer a set of basic queries.	6	WS	PR	Intro	ENG
DMH Data Management on Modern Hardware	Learn the fundamentals of cache-efficient storage and processing models and the basics of parallel data processing on modern CPUs and co-processors for typical database operators.	6	SS	IV	ADV	ENG
IP4RS Image Processing for Remote Sensing	Gain theoretical and practical knowledge on fundamental concepts and techniques for processing and analysis of remote sensing images acquired by Earth observation satellite and airborne systems.	6	SS	IV	ADV	ENG
MDS Management of Data Streams	Develop deep skills in conventional, methodical and the practical processing of continuous data streams using various application examples.	6	WS	IV	ADV	ENG
AMLS Architecture of Machine Learning Systems	Learn about the architecture and essential concepts of modern ML systems for both local and large-scale machine learning, including systems for data-parallel execution, parameter servers, ML lifecycle systems, and the integration of ML into database systems.	6	SS	VL + UE	ADV	ENG
DIA Data Integration and Large-Scale Analysis	Learn about major data integration architectures, key techniques for data integration and cleaning, as well as methods for large-scale, i.e., distributed, data storage and analysis.	6	WS	VL + UE	ADV	ENG
EDML Engineering Data for Machine Learning	This highly technical course focuses on the engineering and life-cycle management of data for production machine learning deployments.	6	WS	VL + UE	ADV	ENG
Data Integration: Algorithms and Systems	Learn about the entire pipeline of an information integration workflow, by learning about existing integration architectures, algorithms in data cleansing, schema matching, and data fusion.	6	WS	VL + TUT	ADV	ENG
BDSPRO Big Data Systems Project	Solve a current research problem in the field of technologies and systems for big data analytics / data science.	9	WS/SS	PJ	ADV	ENG
CV4RS Project Computer Vision for Remote Sensing	Gain practical experience in applying computer vision techniques to address Earth observation questions in a collaborative team and acquire knowledge on state-of-the-art topics in the field of computer vision for remote sensing.	9	WS/SS	PJ	ADV	ENG
Master Project: Large Scale Data Integration	Develop solutions for large scale data integration by working in groups.	9	WS	PJ	ADV	ENG
LDE Large-Scale Data Engineering	In this combined seminar/project module, you will learn about scientific reading and writing and create prototypes of programming projects in data and ML systems, in the context of big data engineering.	12	WS/SS	PJ + SE	ADV	ENG
BDASEM Big Data Analytics Seminar	Learn how to critically read state-of-the-art publications on technologies and systems for big data management and data science as well as learn how to offer an effective presentation, all under the guidance of an assigned mentor.	3	WS/SS	SE	ADV	ENG
Data Integration Seminar	Learn about the most recent trends and challenges in the field of data integration.	3	WS/SS	SE	ADV	ENG
IMSEM Seminar Hot Topics in Information Management	Learn how to critically read state-of-the-art publications on technologies and systems for big data management and data science. In addition, learn how to offer an effective presentation and write a scientific/technical report, all under the guidance of an assigned mentor.	3	WS/SS	SE	ADV	ENG
ML4RS Machine Learning for Remote Sensing Data Analysis	Learn about the current developments in remote sensing and related data analysis methods, and how machine learning techniques can be employed to solve Earth observation questions.	3	WS	SE	ADV	ENG
MLDMS Joint Seminar on Machine Learning and Data Management Systems	In this research-oriented seminar you will learn about and present a selected machine learning and data management topic.	3	WS/SS	SE	ADV	ENG
RDSEM Seminar on Responsible Data Engineering	Learn how to critically read and interpret scientific papers on responsible data management and responsible data science, how to give a good scientific presentation and, how to write a scientific survey. Further, you will learn about state-of-the-art and current research topics in responsible data science and data management.	3	WS	SE	ADV	ENG
DL4CV Recent Trends in Deep Learning for Computer Vision	Acquire advanced knowledge in the field of deep learning for computer vision. Learn how to conduct a literature search, to contextualize scientific papers and to effectively present a scientific topic in an oral presentation/technical report.	3	SS	SE	ADV	ENG
ROC Foundations for Graduate Research in Data Management and Machine Learning Systems	Learn contemporary research methodology, gain both theoretical and practical skills in data management and big data technologies, and be attuned to today's major research challenges in scalable data management and processing.	6	WS	PR + SE	ADV	ENG
APP4RS Advanced Python Programming for Deep Learning in Remote Sensing	Acquire advanced knowledge in Python programming in the field of deep learning for remote sensing, learn how to overcome current remote sensing data engineering challenges, and gain practical skills in training current deep learning methods.	9	WS/SS	PR	ADV	ENG
MTC Master's Thesis Colloquium in Data Management Systems	Covers a presentation on open thesis topics at DIMA, a presentation and discussion on the expectations and evaluation criteria of Master's Theses, topic selection, structuring and writing a thesis proposal, including problem statement, solution approach, experimental design, evaluation, and implementation plan as well as project and time management over the course of a Master's Thesis.	3	WS/SS	CO	ADV	ENG
Research Seminar Data Integration and Data Preparation	Learn about a variety of topics at the intersection of databases and machine learning that aim at solving hard integration and data enhancement problems.	3	WS/SS	CO	ADV	ENG
MT Master's Thesis	Create a scientific work that solves a data management, big data engineering, or technologies and systems problem: (i) surveying related work, (ii) stating the research problem, (iii) defining the scope, (iv) specifying a solution approach and methodology, (v) differentiating the solution from the state-of-the-art, and (vi) showing the effectiveness and efficiency of your solution.	30	WS/SS	THESIS	ADV	GER/ENG
RC DIMA Research Colloquium	Learn about the latest research results from prominent academic or industry guest speakers.	0	WS/SS	CO	ADV	ENG

PATH 1: JOIN THE DATA SYSTEMS LAB AS A PHD STUDENT

- Experience the research process from idea generation, prototype design and implementation to experimental and analytical evaluation
- Gain deep knowledge in your specialization area
- Identify open research questions, devise novel solutions, and validate them
- Make an impact with your own scientific contributions
- Learn proven techniques to disseminate and publish your findings at top-tier venues (e.g., VLDB, SIGMOD, ICDE, EDBT)
- Contribute to large-scale & open-source software projects
- Identify and critically read leading scientific works
- Possibility for excellent PhD students to participate in Software Campus
- Gain technical expertise in database systems, data analysis, data mining, and related topics
- Enhance business and networking skills & interface with researchers and business leaders
- Gain professional work experience by engaging in summer internships

PATH 2: JOIN OR FOUND A STARTUP

Founded by the original creators of Apache Flink in 2014, dataArtisans sought to offer an innovative large-scale data processing technology rooted in sound database and distributed systems principles and architectures. Acquired by Alibaba in 2019, dataArtisans is known as Ververica today. At dataArtisans, DIMA alumni were employed as a CEO, CTO, or software engineer. ParStream introduced the industry's first fully integrated, tested, fast and low-latency big-

data analytics platform built for the Internet of Things. In 2015, ParStream was acquired by Cisco. At ParStream, DIMA alumni were employed as software engineers.

Comprised of an interdisciplinary team of AI scientists, pathologists, software engineers and business professionals, Aignostics is dedicated to taking explainable AI from their research lab to the world. At Aignostics, a DIMA alumnus is employed as the Chief Data Officer.

PATH 3: JOIN INDUSTRY

Representative examples of positions and employers of Data Systems Lab alumni:

- Applied Scientist
- Software Engineer
- Senior Member Technical Staff
- Postdoctoral Researcher
- Researcher
- Senior Research Scientist

