# The Technische Universität Berlin

## Faculty IV Electrical Engineering and Computer Science

# The Data Science and Engineering (DS&E) Master's Track: A Guidance Document (Version 4.3)

**Dr. Tina Schwabe, Juan Soto, and Prof. Dr. Volker Markl**
Berlin Institute for the Foundations of Learning and Data (BIFOLD)
**Technische Universität Berlin**
**Last Updated:** November 14, 2024

**Synopsis**. The *Data Science & Engineering* (formerly, Data Analytics) *Master's Track*, enables students pursing a M.Sc. in Computer Science, Information Systems Management or Computer Engineering, to specialize in data science and engineering. To meet the track requirements, students must complete courses in three core competencies: (1) *data analytics*, (2) *scalable data management*, and (3) *a domain-specific application area* as well as complete a Master's Thesis in data science or data engineering. This guidance document offers students general advice in the selection of courses, the procedure to follow when identifying a thesis topic, and prospective career possibilities. Students who complete both their respective M.Sc. degree and track requirements, will receive – besides their M.Sc. degree – a *Data Science and Engineering Master's Track Certificate* issued by Faculty IV. ***Questions or comments concerning this document should be directed to tina dot schwabe at tu-berlin dot de***.

## 1. Motivation[1]

The last decades were marked by the digitization of virtually all aspects of our daily lives. Today, industry, government institutions and NGOs, and, of course, science and engineering face an avalanche of digital data daily. Partially due to a reduction in disk storage costs, a paradigm shift towards cloud storage services, and the ubiquitous availability of networked devices. At first glance, this appears to be favorable for our increasingly networked society. However, in many ways it is a burden.

Data (often appearing as 'raw data') is neither information, nor knowledge. Data is of great value, once it has been refined and analyzed, to address well-formulated questions, concerning problems of interest. It is only then that economic and social benefits can be fully realized. Modern big data analytics questions are often solved using techniques drawn from varying fields, including graph and network analysis, machine learning, mathematics, statistics, signal processing, and text processing, among others.

Currently, data scientists, well versed in (scalable) data analysis methods, scalable systems programming, and knowledge in an application domain are needed to derive insight from big data. Unfortunately, data scientists with skills in both scalable systems and (potentially domain specific) data analysis methods are few in number. They are expensive and in high-demand. Consequently, this limits the amount of value that can currently be generated from big data for society as a whole.

Moreover, despite the ever-increasing number of data science programs at universities worldwide and student enrollments, it will still be impossible to educate, so-called *Jack-of-all-trades*, given that the skills required are complex and diverse (as depicted in Figure 1). Prior to the rise of the term *big data*, only a few programmers with MPI expertise, predominantly located in supercomputing centers were sufficient in number. For many decades, software engineers and general users in varying domains did not have to worry about scalability issues in their computing systems, thanks in part to higher-level programming languages, compilers, and database systems. In contrast, today's existing technologies have reached their limits due to big data requirements, which involve data volume, data rate and heterogeneity, and the complexity of the analytics. Indeed, the need for more advanced analytics will go beyond relational algebra. They will need to employ complex user-defined functions and support both iterations and distributed state.

[1] *The motivation section was predominantly drawn from Prof. Volker Markl [1, 2].*

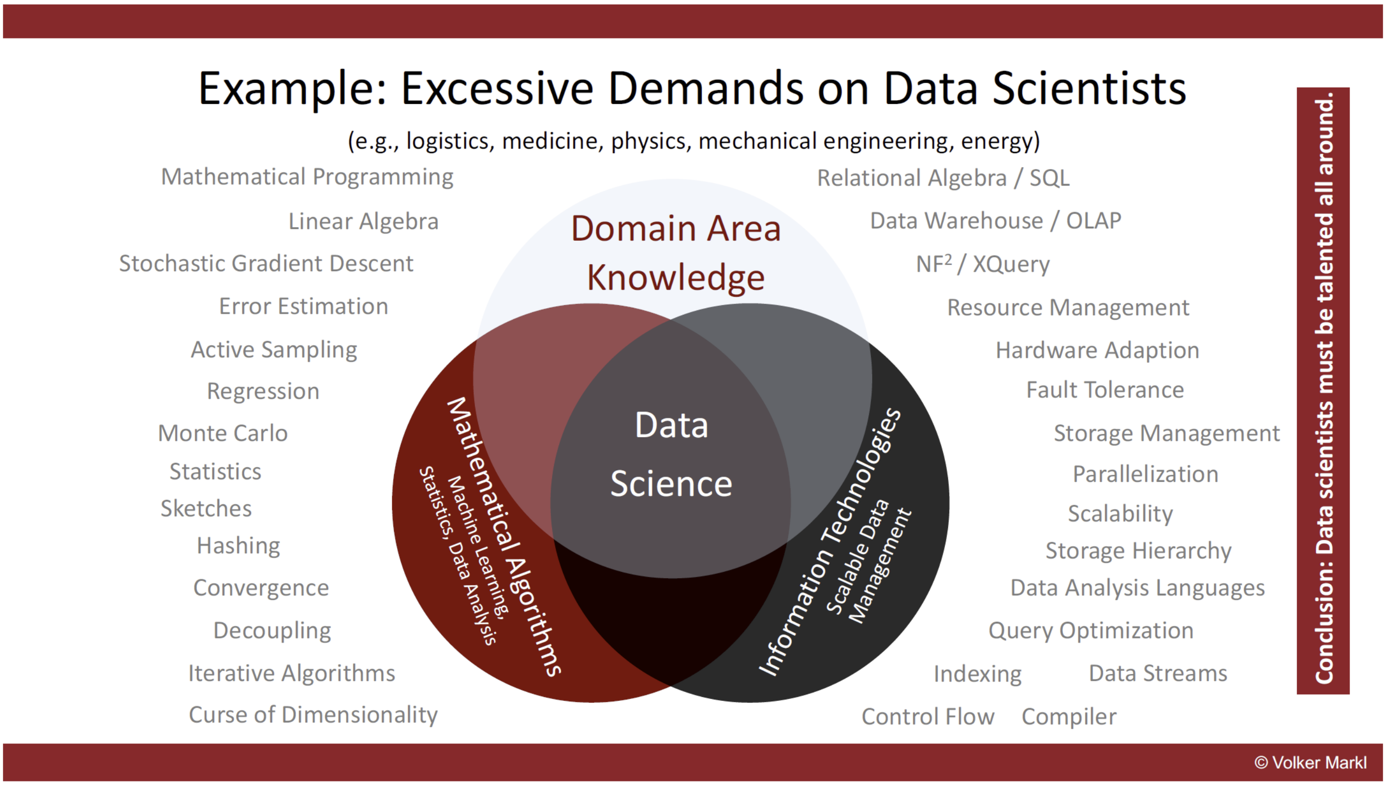


**Figure 1.** The vast array of demands placed on data scientists today.

In the era of many-core processors, cloud computing, and NoSQL, we must ensure that well-established declarative language concepts (inherent in relational database systems) make their way into big data systems. To make this a reality, the research community will need to address the related challenges. For example, (i) designing a programming language specification that does not require systems programming skills, (ii) mapping programs expressed in this programming language to a computing platform of their own choosing, and (iii) executing these in a scalable manner.

This means devising execution strategies that are distributed, parallelized, and support both in-memory technologies and out-of-core execution for data-intensive algorithms. To meet this challenge the compiler, data analysis, database systems, distributed systems, and machine learning communities, among others, will have to come together. We will have to develop novel scalable algorithms and systems that can organize the data deluge and distill information to create value.

Furthermore, the power of declarative languages, to enable *automatic optimization*, *parallelization*, and the *adaptation of a program to varying distributed systems* and *novel hardware architectures* (depending on data distribution, data size, data rate, and system load) must be preserved. In this way, we will overcome the current "stone age" in big data analytics. That is, algorithm specifications in systems that do not automatically optimize (e.g., MPI, MapReduce), imperative languages (e.g., C), object-oriented languages (e.g., Java), and relational-oriented languages (e.g., SQL, XQuery) with non-tunable external driver programs, and technical computing systems (e.g., R, MATLAB) that do not scale.

## 2. Detailed Descriptions of the Master's Track Rules

Please study the following subsections very carefully, most of your questions should be answered.

**2.1 Qualification and Main Competence Areas.** The Data Science and Engineering Master's Track qualifies students to pursue careers as a *Data Scientist*, *Data Analyst*, or *Data Engineer*. They will learn about data analysis methods, their application to real-world problems in varying domains, learn more about the internals of database systems, and develop programming skills with a focus on massively-parallel data processing systems.

**2.2 Requirements.** Students following the track should be enrolled in one of the following TU Berlin Master's Programs: *Computer Science* ('Informatik'), *Information Systems Management* ('Wirtschafts-informatik') or *Computer Engineering* ('Technische Informatik'). Their acceptance to the Data Science and Engineering Track is automatic.

**2.3 Prerequisites.** Students interested in joining the track should possess: (a) very strong English language skills, (b) programming skills in functional (e.g., Scala) and object-oriented (e.g., Java) programming languages, (c) fundamental skills in database management systems, and (d) knowledge in mathematical foundations (e.g., linear algebra, probability, statistics).

**2.4 Credit Points (ECTS) and Track Structure.** To earn a M.Sc. degree, students must achieve 120 ECTS. Of these, 90 ECTS must fulfill the requirements described below, to qualify for the track certificate.

| Credit Points | Competence | Course | Notes[2] |
|---|---|---|---|
| 24 ECTS | Data Analytics (DA) | Machine Learning 1 or Machine Intelligence I | *mandatory* course |
| | | DA Elective 1 | see Appendix A, Table 1 |
| | | DA Elective 2 | |
| | | DA Elective 3 | |
| 18 ECTS | Scalable Data Management (SDM) | Database Technology | *mandatory* course |
| | | SDM Elective 1 | see Appendix A, Table 2 |
| | | SDM Elective 2 | |
| 6 ECTS | Domain Specific Application (DSA) | DSA Elective | see Appendix A, Table 3 |
| 9 ECTS | Project | Project Elective | see Appendix A, Table 4 |
| 3 ECTS | Seminar | Seminar Elective | see Appendix A, Table 5 |
| 30 ECTS | Thesis | Master's Thesis | The thesis must be a *data science-oriented* topic, supervised by a TU Berlin Professor usually from Fak. IV. |
| **Total:** 90 ECTS | | | |

**2.5 Enrolling in the Track.** To enroll in the track, students must join the "*Data Science & Engineering Track*" course located at https://isis.tu-berlin.de/course/view.php?id=40142. Students are advised to complete the Excel spreadsheet available for download from the abovementioned website and forward it on to Tina Schwabe (tina dot schwabe at tu-berlin dot de) for review.

**2.6 Changes to the Track.** Track requirements may change annually. Therefore, students are required to regularly monitor announcements posted on the ISIS *Data Science and Engineering* Track forum.

## Appendix A. Representative List of Master's Courses Across Competence Areas

**Special Instructions (Read Carefully):**

1. Below we list a *representative* list of elective courses that should meet track requirements across varying competencies. Students are advised to complete the Excel spreadsheet available for download from the abovementioned website and forward it on to Tina Schwabe (tina dot schwabe at tu-berlin dot de) for review. If a student wishes to enroll in a course that is not explicitly listed in one of the tables listed below, then you are urged to reach out to *Tina Schwabe* via email or in person, to obtain assurance that the course meets track requirements, **prior to enrolling in the course**.

2. **TU Berlin's course catalog is fairly vast. Thus, we are unable to maintain an accurate record, in this document.** For example, regarding when a course will be offered (i.e., WiSe/WS or SoSe/SS), the specific target language used in class (i.e., EN or DE), or whether new courses will be coming online, among other things. Therefore, students are responsible to obtain the latest information. Students are urged to review the latest course offerings as contained in the Technische Universität Berlin *Course Catalog*: https://moseskonto.tu-berlin.de/moses/modultransfersystem/bolognamodule/suchen.html.

3. Unfortunately, **course schedules (i.e., day and time) are subject to change.** There have been instances where some courses are offered at the exact day and time. In these cases, students should seek to resolve scheduling conflicts by appropriately selecting their courses.

[2] Caveat: Courses listed in the appendices are suggestions. Be aware that some of the existing courses may be removed from the course catalog, while others may be added each term. It is the student's responsibility to request a review of their proposed plan each term.

4. ***Project* / *Seminar* courses can only be applied to the *Project* / *Seminar* requirement, respectively.**

5. ***Data Analytics courses*** should mainly be theory (foundations) courses. A maximum of one practical training course can be chosen.

6. **For a current list of courses students are advised to visit the following groups and their respective webpages.** Courses are primarily drawn from varying research groups in Fak. IV. A representative list is shown below. Note: We are unable to list all of the groups, since the list is dynamic and ever-growing. For an up-to-date list visit: https://www.tu.berlin/eecs/einrichtungen/professuren-fachgebiete.

| Group | Professors |
| --- | --- |
| Agent Technologies in Business Applications & Telecommunication | Prof. Dr. Sahin Albayrak |
| Algorithmics and Computational Complexity | Prof. Dr. Mathias Weller |
| Big Data Engineering | Prof. Dr. Matthias Böhm |
| Communication Systems | Prof. Dr. Thomas Sikora |
| Communications and Information Theory | Prof. Dr. Guiseppe Caire |
| Computer Vision & Remote Sensing | Prof. Dr. Olaf Hellwich |
| Database Systems and Information Management | Prof. Dr. Volker Markl |
| Data Engineering for Machine Learning | Prof. Dr. Sebastian Schelter |
| Data Integration und Data Preparation | Prof. Dr. Ziawasch Abedjan |
| Distributed and Operating Systems | Prof. Dr. Odej Kao |
| Econometrics and Business Statistics | Prof. Dr. Axel Werwatz |
| Efficient Algorithms | N.N. |
| Embedded Systems Architecture | N.N. |
| Image Communication | Prof. Dr. Thomas Wiegand |
| Information Systems Engineering | Prof. Dr. Stefan Tai |
| Intelligent Systems | Prof. Dr. Marc Toussaint |
| Internet and Society | Prof. Dr. Bettina Berendt |
| Internet Architecture and Management | Prof. Dr. Stefan Schmid |
| Language and Communication in Biological and Artificial Systems | Prof. Dr. Fatma Deniz |
| Machine Learning | Prof. Dr. Klaus-Robert Müller |
| Machine Learning and Communication | Prof. Dr. Wojciech Samek |
| Machine Learning and Security | Prof. Dr. Konrad Rieck |
| Scalable Software Systems | Prof. Dr. David Bermbach |
| Modeling of Cognitive Processes | Prof. Dr. Henning Sprekeler |
| Models and Theory of Distributed Systems | Prof. Dr. Uwe Nestmann |
| Network Information Theory | Prof. Dr. Slawomir Stanczak |
| Neural Information Processing | Prof. Dr. Klaus Obermayer |
| Neurotechnology | Prof. Dr. Benjamin Blankertz |
| Open Distributed Systems | Prof. Dr. Manfred Hauswirth |
| Quality and Usability Lab | Prof. Dr. Sebastian Möller |
| Remote Sensing Image Analysis | Prof. Dr. Begüm Demir |
| Robotic Interactive Perception | Prof. Dr. Guillermo Gallego |
| Robotics and Biology Laboratory | Prof. Dr. Oliver Brock |
| Service-centric Networking | Prof. Dr. Axel Küpper |
| Telecommunication Networks | Prof. Dr. Falko Dressler |
| Uncertainty, Inverse Modeling and Machine Learning | Prof. Dr. Stefan Haufe |

**Table 1.** A Representative List of Eligible *Data Analytics* Courses.

| Course Title | Module No. | ECTS | Professor |
|---|---|---|---|
| Machine Learning 2 | 40551 | 9 | Klaus-Robert Müller |
| Machine Learning Lab Course | 40635 | 9 | Klaus-Robert Müller |
| Machine Intelligence II | 40549 | 6 | Klaus Obermayer |
| Machine Learning for Computer Security | 41101 | 6 | Konrad Rieck |
| Adversarial Machine Learning | 41117 | 6 | Konrad Rieck |
| Smart Security Lab | 41116 | 6 | Konrad Rieck |
| Deep Learning 1 | 41071 | 6 | Grégoire Montavon |
| Deep Learning 2 | 41072 | 6 | Klaus-Robert Müller |
| Image Processing for Remote Sensing | 40937 | 6 | Begüm Demir |
| Medical Image Processing | 40882 | 6 | Anja Hennemuth |
| Econometric Analysis of Longitudinal and Panel Data | 70120 | 6 | Axel Werwatz |
| Introduction to Financial Econometrics | 70173 | 6 | Axel Werwatz |
| Microeconometrics | 70187 | 6 | Axel Werwatz |
| Multivariate Analysis/Business Statistics | 70190 | 6 | Axel Werwatz |
| Time Series Analysis | 70250 | 6 | Axel Werwatz |
| Treatment Effect Analysis | 70251 | 6 | Axel Werwatz |
| Ökonometrie (Econometrics) | 70198 | 6 | Axel Werwatz |
| Natural Language Processing | 41047 | 6 | Sebastian Möller |
| Digital Image Processing | 40414 | 6 | Olaf Hellwich |
| Automatic Image Analysis | 40345 | 6 | Olaf Hellwich |
| Applied Environmental Econometrics in R | 70383 | 6 | Astrid Cullmann |
| AI and Cybersecurity | 40900 | 6 | Sahin Albayrak |

**Table 2.** A Representative List of Eligible *Scalable Data Management* Courses.

| Course Title | Module No. | ECTS | Professor |
|---|---|---|---|
| MDS Management of Data Streams | 40310 | 6 | Volker Markl |
| DBTLAB Database Technology Lab | 40037 | 6 | Volker Markl |
| DMH Data Management on Modern Hardware | 40804 | 6 | Volker Markl |
| Architecture of Machine Learning Systems | 41078 | 6 | Matthias Böhm |
| Data Integration and Large-scale Analysis | 41112 | 6 | Matthias Böhm |
| Cloud Computing | 40368 | 6 | Odej Kao |
| Cloud Native Architecture and Engineering | 40103 | 6 | Stefan Tai |
| Algorithms for Distributed Systems | 41127 | 6 | Stefan Schmid |
| EDML - Engineering Data for Machine Learning | 41221 | 6 | Sebastian Schelter |
| Data Integration: Algorithms and Systems | 41213 | 6 | Ziawasch Abedjan |

**Table 3.** A Representative List of Eligible *Domain Specific Application* Courses.

| Course Title | Module No. | ECTS | Professor |
|---|---|---|---|
| Digitale Märkte (Digital Markets) | 70414 | 6 | Nancy Wünderlich |
| Energy Economics - Energy Sector Modeling (EW-MOD) | 70129 | 6 | Christian Hirschhausen |
| Energiewirtschaft - Technologie u. Innovation (EW-TUI) | 70132 | 6 | Christian Hirschhausen |
| Energy Economics | 30024 | 6 | Thomas William Brown |
| Gesundheitsökonomie II (Health Economics) | 70142 | 6 | Marco Runkel |
| Integriertes Informationsmanagement | 70166 | 6 | Rüdiger Zarnekow |
| IT-Service-Management | 70175 | 6 | Rüdiger Zarnekow |
| Patentrecht und Patentmanagement I (Patent Rights and Patent Management) | 70000 | 6 | Martin Sebastian Haase |
| Speech Signal Processing and Speech Technology | 40721 | 6 | Sebastian Möller |
| The Economics of Climate Change | 60431 | 6 | Ottmar Georg Edenhofer |
| Auctions: Theory and Applications | 70373 | 6 | Radosveta Ivanova-Stenzel |
| Psychology for Engineers | 50535 | 6 | Eva Wiese |
| Energie und Ressourcen - Grundlagen | 70125 | 6 | Joachim Müller-Kirchenbauer |
| Introduction to Space Geodesy | 61440 | 6 | Frank Flechtner |
| Global Logistics Management | 70143 | 6 | Frank Straube |

**Table 4.** A Representative List of Eligible *Project* Courses.

| Course Title | Module no. | ECTS | Professor |
|---|---|---|---|
| BDSPRO Big Data Systems Project | 40494 | 9 | Volker Markl |
| ROC Foundations for Graduate Research in Data Management and Machine Learning Systems | 41135 | 9 | Volker Markl |
| Master Project: Distributed Systems | 40552 | 9 | Odej Kao |
| Machine Learning Project | 40653 | 9 | Klaus-Robert Müller |
| Machine Learning and Security - Project | 41102 | 9 | Konrad Rieck |
| Projekt Neuronale Informationsverarbeitung | 40654 | 9 | Klaus Obermayer |
| Projekt Nachrichtenübertragung (Signal Processing Project) | 40161 | 6 | Thomas Sikora |
| Project Large-scale Data Engineering | As part of: 41086 | 9 | Matthias Böhm |
| Project Computer Vision for Remote Sensing | 41012 | 9 | Begüm Demir |
| Internet of Services Lab (Project) | 40514 | 9 | Axel Küpper |
| Data Science Project | 40693 | 9 | Sahin Albayrak |
| Advanced Distributed Systems Prototyping | 40984 | 12 | David Bermbach |
| Advanced Cloud Prototyping | 41153 | 12 | Stefan Tai |
| Master Project: Large Scale Data Integration | 41215 | 9 | Ziawasch Abedjan |

**Table 5.** A Representative List of Eligible *Seminar* Courses.

| Course Title | Module No. | ECTS | Professor |
|---|---|---|---|
| BDASEM Big Data Analytics Seminar | 40353 | 3 | Volker Markl |
| IMSEM Seminar on Hot Topics in Information Management | 40001 | 3 | Volker Markl |
| Seminar Large-scale Data Engineering | As part of: 41086 | 3 | Matthias Böhm |
| Machine Learning and Data Management Systems | 41146 | 3 | Matthias Böhm |
| Machine Learning and Security – Master Seminar | 41104 | 3 | Konrad Rieck |
| Machine Learning in Science and Industry | 41044 | 3 | Grégoire Montavon |
| Machine Learning for Remote Sensing Data Analysis | 40928 | 3 | Begüm Demir |
| Internet of Services Lab (Seminar) | 41043 | 3 | Axel Küpper |
| Master Seminar: Operating Complex IT Systems | 40036 | 3 | Odej Kao |
| Uncertainty in Machine Learning | 41113 | 3 | Stefan Haufe |
| Ethics, Data Science, and Networked AI | Part of 40994 | 3 | Bettina Berendt |
| Seminar Hot Topics in Computer Vision | 40488 | 3 | Olaf Hellwich |
| Machine and Behavior | 41184 | 3 | Oliver Brock |
| RDSEM - Seminar on Responsible Data Engineering | 41224 | 3 | Sebastian Schelter |
| Data Integration Seminar | 41214 | 3 | Ziawasch Abedjan |

## Appendix C. Frequently Asked Questions

### Q1. What is a track?

**A1.** In general, a track is a suggested sequence of courses that profile a specific specialization. Students who successfully complete the track will be awarded a certificate from Faculty IV. A certificate indicates that a student has followed a structured academic program with the intent to pursue specialization in data science.

### Q2. Who can follow a track?

**A2.** By default, students enrolled in the Computer Science ("*Informatik*"), Information Systems Management ("*Wirtschaftsinformatik*") or Computer Engineering ("Technische Informatik") Master's programs are eligible to pursue the track. **Unfortunately, due to resource constraints, we are unable to consider other study programs at this time beyond the three mentioned above.**

### Q3. Will my study period be extended, if I follow the track?

**A3.** No, neither the amount of ECTS credit points, nor the number of semesters will increase. Moreover, a longer study period will not lead to a disqualification from the track.

### Q4. How to go about selecting a thesis topic?

**A4.** Students should speak with Senior Researchers, Postdocs, or PhD students, in the participating research groups, i.e. "Chairs," to identify an open thesis topic of mutual interest. For a list of representative data science-oriented publications have a look at [3, 4, 5], and for Master's Thesis topics see [6]. For a glimpse into ongoing research activities in big data/data science see [7]. For open problems and a vision of the future of computer science see [8, 9], respectively. For further discussion about the evolution of the field and varying applications across Germany see [10, 11], respectively.

**Q5. What are my prospective career possibilities?**

**A5.** Students who complete the data analytics track are prepared to pursue careers as *Data Analysts*, *Data Engineers*, or *Data Scientists*. For information about big data projects in industry within Germany have a look at [12]. In some cases, students enter a PhD program with the aim to further specialize in a research topic, such as *deep learning* or *streaming systems*. Examples of recent (DIMA specific) PhD thesis topics, include [13, 14, 15, 16, 17, 18]. Recent (ML/IDA specific) PhD thesis topics, include [19, 20]. For more information about job opportunities and earning potential across Europe have a look at [21, 22].

**Q6. If I still have questions or doubts, not answered yet?**

**A6.** This document is assumed to be comprehensive. It should address the most relevant questions. In case of any doubt (e.g., you are enrolled in a different study programme) or concern, please contact Dr. Tina Schwabe (tina dot schwabe at tu-berlin dot de). Also, please look for announcements (e.g., the bi-annual "*Data Science and Engineering Track Intro Presentation*") posted on the *Data Science and Engineering* Track forum in ISIS.

**Q7. How do I obtain my certificate?**

**A7.** You will need to present evidence (e.g., academic transcript) that you have met the track requirements. Once this has been verified, we will prepare your certificate.

## Appendix D. Version History

| Version | Authors | Date | Remarks |
|---|---|---|---|
| 1.1 | M. Schubotz, H. Hemsen, V. Markl | 28.06.13 | Initial version in German |
| 1.2 | M. Schubotz, J. Soto, V. Markl | 31.07.15 | Translation into English |
| 1.3 | M. Schubotz, J. Soto, V. Markl | 16.01.16 | Updates and Revisions |
| 2.0 | R. Kutsche, V. Markl, J. Soto | 09.10.17 | Full Revision, new version 2 |
| 3.0 | R. Kutsche, V. Markl, J. Soto | 05.03.19 | Track name change, clarification on course selection. |
| 4.0 | V. Markl, J. Soto | 07.10.20 | Removal of courses that are no longer offered, replacement of broken links, removal of sample curriculum, insertion of the URLSs corresponding to the teaching webpages for varying university groups. |
| 4.1 | V. Markl, J. Soto | 14.10.20 | Revision of Q2 to limit the track to: CS, CE, ISM. |
| 4.2 | V. Markl, J. Soto, T. Schwabe | 31.03.24 | Overall revision of the whole document, e.g., App. A, 5. new, new person of contact, list of varying university groups updated (URLs where deleted; App. A, 6.), addition of new courses and removal of courses that are no longer offered (App. A, tables 1-5), replacement and updating of broken links. |
| 4.3 | V. Markl, J. Soto, T. Schwabe | 14.11.24 | Updates and revisions |

## References

[1] "*Breaking the Chains: On Declarative Data Analysis and Data Independence in the Big Data Era*," Volker Markl, PVLDB, 7(13):1730–1733, 2014. URL: www.vldb.org/pvldb/vol7/p1730-markl.pdf.

[2] *Towards a Thriving Data Economy: Open Data, Big Data, and Ecosystems* (Presentation), Volker Markl, European Competitiveness Council, March 2015. URL: https://tubcloud.tu-berlin.de/s/DFGtc3Kmgsk2Hwc.

[3] FG DIMA Data Science Publications. URL: https://www.tu.berlin/en/dima/research/publications.

[4] FG DAMS Data Science Publications. URL: https://www.tu.berlin/en/dams/research/publications.

[5] FG ML/IDA Machine Learning Publications. URL: https://web.ml.tu-berlin.de/publication/.

[6] *Completed Master's Theses*: Many Data Science Oriented, DIMA Group.
URL: https://www.tu.berlin/en/dima/teaching/thesis-opportunities/completed-bachelor-and-master-theses.

[7] Berlin Institute for the Foundations of Learning and Data (BIFOLD). URL: https://bifold.berlin/.

[8] *50 Years of Data Science* (Version 1.00), David Donoho, Stanford University, September 2015.
URL: http://courses.csail.mit.edu/18.337/2015/docs/50YearsDataScience.pdf.

[9] Frontiers in Massive Data Analysis, National Academies Press, 2013. URL: http://nap.edu/18374.

[10] 2021 Data/AI Salary Survey, Mike Loukides, O'Reilly Press, 2021.
URL: https://www.oreilly.com/radar/2021-data-ai-salary-survey.

[11] "Data Science–A Systematic Treatment," CACM, July 2023.
URL: https://www.youtube.com/watch?v=m9XecEc9yGw

[12] Germany – Excellence in Big Data, Bitkom, 2016.
URL: https://www.bitkom.org/Bitkom/Publikationen/Germany-Excellence-in-Big-Data.html.

[13] *Scaling Data Mining in Massively Parallel Dataflow Systems* (PhD Thesis), S. Schelter, November 2015.

[14] *Visualization-Driven Data Aggregation* (PhD Thesis), U. Jugel, TU Berlin, April 2017.

[15] *Adaptive Parameter-Server* (PhD Thesis), A. Renz-Wieland, TU Berlin, December 2022.

[16] *Accelerating Approximate Data Analysis with Parallel Processors* (PhD Thesis), M. Kiefer, TU Berlin, February 2023.

[17] *Query Compilation for Modern Data Processing Environments* (PhD Thesis), P. Grulich, TU Berlin, November 2023.

[18] *Query Processing on Heterogeneous Systems* (PhD Thesis), V. Rosenfeld, TU Berlin, December 2023.

[19] *XAI for Unsupervised Learning (PhD Thesis),* J. R. Kauffmann, TU Berlin, December 2023.

[20] *Debugging learning algorithms: Understanding and correcting machine learning models (PhD Thesis),* C. J. Anders, TU Berlin, January 2024.

[21] *The European Data Science Salary Survey: Tools, Trends, What Pays (and What Doesn't) for Data Professionals in Europe*, John King & Roger Magoulas, O'Reilly Press, 2017.

[22] Plattform Lernende Systeme (Learning Systems Platform), 2024.
URL: https://www.plattform-lernende-systeme.de/home-en.html.