The Technische Universität Berlin

Faculty IV Electrical Engineering and Computer Science

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

Juan Soto, Dr. Ralf Detlef Kutsche, and Prof. Dr. Volker Markl
Database Systems and Information Management (DIMA) Group
Last Updated: March 5, 2019
Email: lehre@dima.tu-berlin.de

Synopsis. In Fall 2013, TU Berlin's Faculty IV Electrical Engineering and Computer Science (EECS) approved a new track, which enables students pursing a M.Sc. in Computer Science, Information Systems Management or Computer Engineering, to specialize in data analytics. To meet the track requirements, students must complete courses in three core competencies: (1) *scalable data analytics*, (2) *scalable data management*, and (3) *a domain-specific application area*. 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. **In April 2019, the track was renamed, the *Data Science & Engineering Master's Track*.** Consequently, from SS 2019 on, students who complete both their respective M.Sc. degree and track requirements, will receive – besides their M.Sc. degree – a *Data Science & Engineering Master's Track Certificate*, issued by Faculty IV.

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 Figures 1 and 2). 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

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

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.

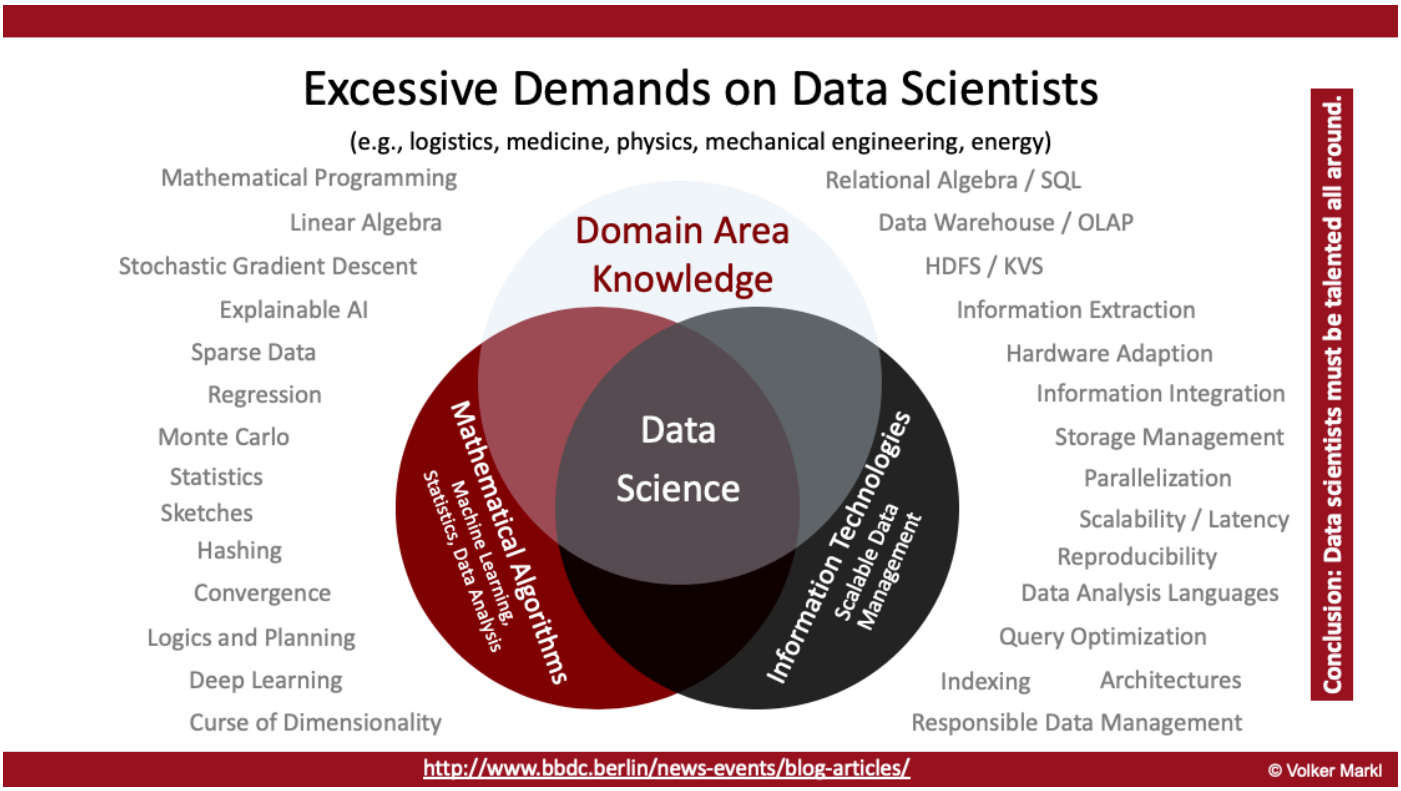

Figure 1. The vast array of demands placed on data scientists today.

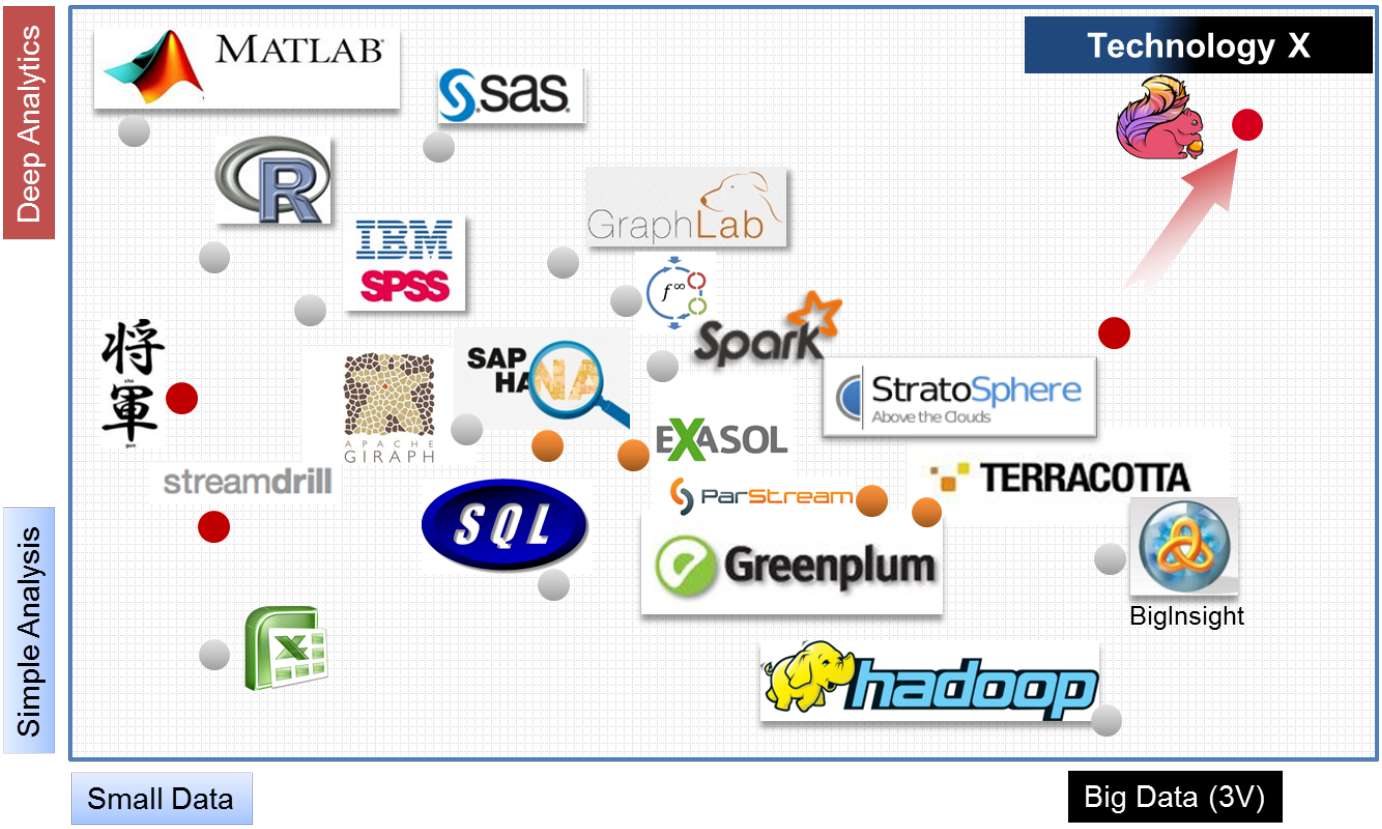

Figure 2. Deep analysis is the name of the game.

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 Data Science & Engineering 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 Analytics track is by default.* However, very dedicated students from other programmes (e.g. *Mathematics*) also could be eligible. This will be decided on individual request, carefully checking the application and the background from the CV, and from your (in this case: mandatory) motivation letter.

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 and Track Structure. To earn a M.Sc. degree, students must achieve 120 ECTS credit points. Of these, 90 ECTS credit points must fulfill the requirements described further below, to quality for the track certificate.

Credit Points	Competence	Course	Notes
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 Data Analytics Lab Professor.
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://www.isis.tu-berlin.de/2.0/course/view.php?id=1069.

2.6 Mentoring Program. Track participants are invited to contact a member of the Data Analytics Lab to identify a mentor and request guidance.

2.7 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 Elective Master's Courses Across Competency Areas

Special Instructions (Read Carefully):

1. Below we list a *representative* list of elective courses that should meet track requirements across varying competencies. 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 *Dr. Ralf Detlef Kutsche* or *Juan Soto* 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, in this document, we are unable to maintain an accurate record. For example, regarding when a course will be offered (i.e., WiSe or SS), the specific target language spoken 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.

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.

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

Title	ECTS	Prof	Term	Lang
Machine Learning 2	9	Klaus-Robert Müller	Summer	EN + DE
Praktikum Maschinelles Lernen (Machine Learning Lab)	9	Klaus-Robert Müller	Summer	EN
Machine Intelligence II	6	Klaus Obermayer	Summer	EN
Monte Carlo Methods in Machine Learning and AI	6	Manfred Opper	Summer	EN
Probabilistic and Bayesian Modelling in ML and AI	6	Manfred Opper	Summer	EN
Econometric Analysis of Longitudinal and Panel Data	6	Axel Werwatz	Winter	EN
Microeconometrics	6	Axel Werwatz	Winter	EN
Multivariate Analysis/Business Statistics	6	Axel Werwatz	Summer	EN
Time Series Analysis	6	Axel Werwatz	Winter	EN
Treatment Effect Analysis	6	Axel Werwatz	Summer	EN
Ökonometrie (Econometrics)	6	Axel Werwatz	Winter	EN
Numerische Mathematik für Ingenieure II	10	Jörg Liesen	Winter	DE
Stochastische Modelle (Stochastic Models)	10	Michael Scheutzow	Winter	DE
Digitale Signalverarbeitung (Digital Signal Processing)	12	Reinhold Orglmeister	Winter	DE

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

Title	ECTS	Prof	Term	Lang
AIM-1 Heterogeneous & Distributed Info. Systems	6	Ralf-Detlef Kutsche	Winter	EN
AIM-2 Management of Data Streams	6	Volker Markl	Winter or Summer	DE
AIM-3 Scalable Data Science: Systems & Methods (SDSSM)	6	Volker Markl	Winter or Summer	EN
IDB-PRA: Implementation of a Database Engine	6	Volker Markl	Winter	EN
CIT 9 - Cloud Computing	6	Odej Kao	Winter or Summer	EN
Parallel Systems	6	Hans-Ulrich Heiss	Summer	EN

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

Title	ECTS	Prof	Term	Lang
Digital Communities	6	Axel Küpper	Winter	EN
Energiewirtschaft - Elektrizitätswirtschaft	6	Christian Hirschhausen	Summer	DE
Energiewirtschaft - Technologie und Innovation	6	Christian Hirschhausen	Summer	DE
Energy Economics	6	Georg Erdmann	Winter	EN
Experimental and Behavioral Economics	6	Dorothea Kübler	Summer	EN
Gesundheitsökonomie II	6	Marco Runkel	Winter	DE
Integriertes Informationsmanagement	6	Rüdiger Zarnekow	Summer	DE
IT-Service-Management	6	Rüdiger Zarnekow	Winter	DE
Intelligente Sicherheit in Netzwerken (IT Sec. in Networks)	9	Sahin Albayrak	Summer	DE
Patentrecht/Patentmanagement (Patent Rights / Mgmt.)	6	Jürgen Ensthaler	Summer	DE
Signalverarbeitung (Signal Processing)	6	Reinhold Orglmeister	Winter	DE
Speech Signal Processing and Speech Technology	6	Sebastian Möller	Winter	EN + DE
The Economics of Climate Change	6	Ottmar Edenhofer	Summer	EN

Table 3. A representative list of eligible *domain specific application* courses.

Title	ECTS	Prof	Term	Lang
IMPRO3 - Big Data Analytics Project (BDAPRO)	9	Volker Markl	Winter or Summer	EN
Verteilte Systeme (Distributed Systems Project)	9	Odej Kao	Winter	DE
Project Machine Learning	9	Klaus-Robert Müller	Winter	EN
Project Neural Information Processing	9	Klaus Obermayer	Summer	EN + DE
Project: Statistical Methods in AI and ML	9	Manfred Opper	Winter	EN + DE
Projekt Nachrichtenübertragung (Signal Processing Project)	6	Thomas Sikora	Winter or Summer	DE

Table 4. A representative list of eligible *project* courses.

Title	ECTS	Prof	Term	Lang
Anwendungen Kognitiver Algorithmen (Applied Cognitive Algorithms)	3	Klaus-Robert Müller	Winter or Summer	DE
BDASEM - Big Data Analytics Seminar	3	Volker Markl	Winter	EN
CIT 8 - Aktuelle Themen aus dem Bereich der verteilten Systeme (Hot Topics in Distributed Systems)	3	Odej Kao	Winter	DE
Hot Topics in Operating Systems & Distributed Systems	3	Hans-Ulrich Heiß	Winter or Summer	EN
IMSEM - Seminar Hot Topics in Info. Management	3	Volker Markl	Summer	EN
Introduction to Computational Genomics	3	Manfred Opper	Summer	EN
Seminar: Operating Complex IT Systems	3	Odej Kao	Winter or Summer	EN + DE
Recent Advances in Computer Architecture	3	Bernardus Juurlink	Winter	EN
Recent Advances in Multicore Systems	3	Bernardus Juurlink	Summer	EN
Synchronous & Asynchronous Interactions in Distributed Systems	3	Uwe Nestmann	Summer	EN + DE

Table 5. A representative list of eligible *seminar* courses.

Appendix B. Representative Examples of Student Curricula*

1. A Computer Science (Stats/Econometrics Concentration) Sample Curriculum.

First Semester (Winter Term)

Title	ECTS	Area	Term	Lang
Machine Intelligence I or Machine Learning I	6	Data Analytics (mandatory)	Winter	EN
Database Technology	6	Scalable Data Mgmt. (mandatory)	Winter	EN
e.g., IDB-PRA: Implementation of a Database Engine	6	Scalable Data Mgmt.	Winter	EN
e.g., Digital Communities	6	Application	Winter	EN
e.g., BDASEM - Big Data Analytics Seminar	3	Seminar	Winter	EN
Elective	3	(e.g., DA, SDM)	Winter	EN + DE

Second Semester (Summer Term)

Title	ECTS	Area	Term	Lang
e.g., Multivariate Analysis/Business Statistics	6	Data Analytics (DA)	Summer	EN
e.g., Treatment Effect Analysis	6	Data Analytics	Summer	EN
e.g., Machine Intelligence II	6	Data Analytics	Summer	EN
e.g., AIM-3 Scalable Data Science	6	Scalable Data Mgmt. (SDM)	Summer	EN
Elective	6	(e.g., DA, SDM)	Summer	EN + DE

Third Semester (Winter Term)

Title	ECTS	Area	Term	Lang
e.g., Time Series Analysis	6	Data Analytics (DA)	Winter	EN
e.g., Microeconometrics	6	Data Analytics	Winter	EN
e.g., AIM-2 Management of Data Streams	6	Scalable Data Mgmt. (SDM)	Winter	DE
e.g., IMPRO3 - Big Data Analytics Project (BDAPRO)	9	Project	Winter	EN
Elective	3	(e.g., DA, SDM)	Winter	EN + DE

Fourth Semester (Summer Term)

Title	ECTS	Area	Term	Lang
Master's Thesis in the Data Analytics Lab	30	Data Science Oriented	Summer	EN + DE

2. A Computer Science (Mathematics Concentration) Sample Curriculum.

First Semester (Winter Term)

Title	ECTS	Area	Term	Lang
Machine Intelligence I or Machine Learning I	6	Data Analytics (mandatory)	Winter	EN
Database Technology	6	Scalable Data Mgmt. (mandatory)	Winter	EN
e.g., AIM-2 Management of Data Streams	6	Scalable Data Mgmt.	Winter	EN + DE
e.g., IDB-PRA: Implementation of a Database Engine	6	Scalable Data Mgmt.	Winter	EN
e.g., Digital Communities	6	Application	Winter	EN

Second Semester (Summer Term)

Title	ECTS	Area	Term	Lang
e.g., Machine Intelligence II	6	Data Analytics	Summer	EN
e.g., AIM-3 Scalable Data Science	6	Scalable Data Mgmt. (SDM)	Summer	EN
Electives	18	(e.g., DA, SDM)	Summer	EN + DE

Third Semester (Winter Term)

Title	ECTS	Area	Term	Lang
e.g., BDASEM - Big Data Analytics Seminar	3	Seminar	Winter	EN
e.g., IMPRO3 - Big Data Analytics Project (BDAPRO)	9	Project	Winter	EN
Electives	18	(e.g., DA, SDM)	Winter	EN + DE

Fourth Semester (Summer Term)

Title	ECTS	Area	Term	Lang
Master's Thesis in the Data Analytics Lab	30	Data Science Oriented	Summer	EN + DE

3. An Information Systems Management Sample Curriculum.

First Semester (Winter Term)

Title	ECTS	Area	Term	Lang
Machine Intelligence I or Machine Learning I	6	Data Analytics (mandatory)	Winter	EN
Database Technology	6	Scalable Data Mgmt. (mandatory)	Winter	EN
e.g., Multivariate Analysis/Business Statistics	6	Data Analytics (DA)	Winter	EN
e.g., Digital Communities	6	Application	Winter	EN
Elective	6	(e.g., DA, SDM)	Winter	EN

Second Semester (Summer Term)

Title	ECTS	Area	Term	Lang
e.g., AIM-3 Scalable Data Science	6	Scalable Data Mgmt. (SDM)	Summer	EN
e.g., AIM-2 Management of Data Streams	6	Scalable Data Mgmt. (SDM)	Summer	DE
e.g., CIT 9 - Cloud Computing	6	Scalable Data Mgmt.	Summer	EN
e.g., Application System Project[2]	12	Project	Summer	EN

Third Semester (Winter Term)

Title	ECTS	Area	Term	Lang
e.g., Time Series Analysis	6	Data Analytics	Winter	EN
e.g., Microeconometrics	6	Data Analytics	Winter	EN
e.g., IMPRO3 - Big Data Analytics Project (BDAPRO)	9	Project	Winter	EN
e.g., IMSEM - Seminar Hot Topics in Info. Management	3	Seminar	Winter	EN
Elective	6	Data Analytics	Winter	EN

Fourth Semester (Summer Term)

Title	ECTS	Area	Term	Lang
Master's Thesis in the Data Analytics Lab	30	Data Science Oriented	Summer	EN + DE

[2] The particular project undertaken must have data science elements. Otherwise, an alternative project will need to be select.

4. A Computer Engineering Sample Curriculum.

First Semester (Winter Term)

Title	ECTS	Area	Term	Lang
Machine Intelligence I or Machine Learning I	6	Data Analytics (mandatory)	Winter	EN
Database Technology	6	Scalable Data Mgmt. (mandatory)	Winter	EN
e.g., Hot Topics in Next Generation Networks and Future Internet Technologies	3	Seminar	Winter	EN + DE
Electives	15	(e.g., DA, SDM)	Winter	EN + DE

Second Semester (Summer Term)

Title	ECTS	Area	Term	Lang
e.g., Machine Intelligence II	6	Data Analytics	Summer	EN
e.g., Probabilistic and Bayesian Modelling in ML and AI	6	Data Analytics	Summer	EN
e.g. Digitale Signalverarbeitung (Digital Signal Processing)	9	Application	Summer	DE
Electives	9	(e.g., DA, SDM)	Summer	EN + DE

Third Semester (Winter Term)

Title	ECTS	Area	Term	Lang
e.g., IDB-PRA: Implementation of a Database Engine	6	Scalable Data Mgmt.	Winter	EN
e.g., Speech Signal Processing and Speech Technology	6	Application	Winter	EN + DE
e.g., Verteilte Systeme (Distributed Systems Project)	9	Project	Winter	DE
Electives	9	(e.g., DA, SDM)	Winter	EN + DE

Fourth Semester (Summer Term)

Title	ECTS	Area	Term	Lang
Master's Thesis in the Data Analytics Lab	30	Data Science Oriented	Summer	EN + DE

Appendix C. Questions and Answers

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. Students from other study programs should contact the "*Data Analytics Track Steering Board*" at lehre@dima.tu-berlin.de to determine whether they can participate in the track.

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], and for Master's Thesis topics see [5]. For a glimpse into ongoing research activities in big data/data science see [6]. For open problems and a vision of the future of computer science see [7, 8, 9], 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 [10]. 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 [11, 12, 13, 14, 15]. For more info. about job opportunities and earning potential across Europe have a look at [16].

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 us at lehre@dima.tu-berlin.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, DIMA staff will prepare your certificate.

Appendix D. Version History

Version	Authors	Date	Remarks
1.1	Moritz Schubotz, Holmer Hemsen, Volker Markl	28.06.13	Initial version in German
1.2	Moritz Schubotz, Juan Soto, Volker Markl	31.07.15	Translation into English
1.3	Moritz Schubotz, Juan Soto, Volker Markl	16.01.16	Updates and Revisions
2.0	Ralf Detlef Kutsche, Volker Markl, Juan Soto	09.10.17	Full Revision, new version 2
3.0	Ralf Detlef Kutsche, Volker Markl, Juan Soto	05.03.19	Track name change, clarification on course selection.

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: goo.gl/eDRSS3.

[3] DIMA web: Data Science Publications: http://www.dima.tu-berlin.de/menue/publications/publications/.

[4] ML/IDA web: Machine Learning Publications: http://doc.ml.tu-berlin.de/publications/.

[5] *Completed Master's Theses*: Many Data Science Oriented, DIMA Group. URL: http://www.dima.tu-berlin.de/menue/theses/completed_mastersdiploma_theses/.

[6] **Berlin Big Data Center**, http://www.bbdc.berlin/home/.

[7] *Future Directions in Computer Science Research* (Presentation: TU Berlin, Big Data Workshop), John Hopcroft, Cornell University, September 2013. URL: http://www.eecs.tu-berlin.de/index.php?id=139969.

[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] Germany – Excellence in Big Data, Bitkom, 2016. URL: goo.gl/wUSZWv.

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

[12] *Specification and Optimization of Analytical Data Flows* (PhD Thesis), F. Hüske, December 2015.

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

[14] *Exploratory Relation Extraction in Large Multilingual Data* (PhD Thesis), A. Akbik, April 2016.

[15] *Programming Abstractions, Compilation, and Execution Techniques for Massively Parallel Data Analysis* (PhD Thesis), S. Ewen, TU Berlin, November 2014.

[16] *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.