



# Studiengangsdokumentation

MASTERSTUDIENGANG

DATA ENGINEERING AND ANALYTICS

Fakultät für Informatik, Technische Universität München

14. Februar 2019

**Bezeichnung:** Data Engineering und Analytics

**Organisatorische**

**Zuordnung:** Fakultät für Informatik

**Abschluss:** Master of Science (M.Sc.)

**Regelstudienzeit**

**(Credits) (SWS):** 4 Semester (120 CP) (61 SWS)

**Studienform:** Vollzeit

**Zulassung:** Eignungsverfahren

**Starttermin:** Wintersemester 2016/2017

**Sprache:** Englisch, einzelne Module in Deutsch (komplett in Englisch studierbar)

**Studiengangs-**

**verantwortlicher:** Prof. Dr. Thomas Neumann

**Ergänzende Angaben für besondere Studiengänge:** Enge Zusammenarbeit mit dem Studiengang der Fakultät für Mathematik "Mathematics in Data Science"

**Ansprechperson(en) bei**

**Rückfragen:** Prof. Dr. Thomas Huckle

**Version/Stand vom:** 14. Februar 2019

**Der Studiendekan, Prof.  
Dr. Helmut Seidl:**

**Vormerkung zum Sprachgebrauch:**

Nach Art. 3 Abs. 2 des Grundgesetzes sind Männer und Frauen gleichberechtigt. Alle maskulinen Personen- und Funktionenbezeichnungen in dieser Studiengangsdokumentation gelten daher für Frauen und Männer in gleicher Weise.

# 1 Studiengangsziele

## 1.1 Zweck des Studiengangs

Die Verarbeitung und Auswertung von extrem großen Datenmengen ist ein drängendes Problem in vielen Bereichen und erfordert ganz neue Techniken und Verfahren. Dieser Trend zu "Big Data" wird durch eine Reihe von Entwicklungen vorangetrieben: Erstens wird es immer einfacher, große Mengen von Daten zu erzeugen und zu speichern, z.B. wegen günstigem Speicherplatz, Sensoren, *smart devices*, Sozialen Netzwerken, usw. Zweitens können diese großen Datenmengen durch technische Fortschritte wie z.B. schnelle multi-core Systeme und *Cloud Computing* auch effizient verarbeitet werden. Und drittens werden solche Datenmengen nicht mehr nur von "klassischen" Anwendungsgebieten wie z.B. Geschäftsdaten erzeugt, sondern fallen inzwischen in sehr vielen Lebensbereichen an. Beispielsweise erzeugen immer mehr Fahrzeuge mittels Sensoren und intelligenter Vernetzung große Datenmengen, welche u.a. für die Modellweiterentwicklung oder für diagnostische Zwecke genutzt werden können. Ein wichtiges Anwendungsgebiet ist derzeit auch die intelligente Energienetzsteuerung (Stichwort ‚Smart Grids‘).

Die Nutzung von Big Data hat sowohl wirtschaftlichen als auch wissenschaftlichen Nutzen. Klimaforschung, Astrophysik, Medizin, ... – die Liste der Disziplinen, die durch die intelligente Nutzung der großen verfügbaren Datenmengen Fortschritte erzielen und weitere erhoffen, ist lang. In einer Studie des Fraunhofer Instituts für System- und Innovationsforschung (ISI) über „Gesamtwirtschaftliche Potenziale intelligenter Netze in Deutschland“ von 2012 wird ein gesellschaftlicher Gesamtnutzen von 55,7 Mrd. € pro Jahr erwartet. Der Gesamtnutzen setzt sich zusammen aus erwarteten Effizienzsteigerungen (39,0 Mrd. Euro) und zusätzlichen Wachstumsimpulsen (16,7 Mrd. Euro). Gleichzeitig wächst das Bewusstsein in der Gesellschaft - Personen und Organisationen - um den Wert der Daten und sie sorgen sich um deren Schutz. Der Umgang mit den Enthüllungen rund um die NSA, aber auch Sorgen z.B. um mangelnde Kontrolle im Finanzwesen, führen zu einer Debatte über Datenschutz und Datensicherheit, für die auch technische Expertise gefordert ist.

Diese Kombination aus technologischem Fortschritt und sozialer Entwicklung stellt die Datenverarbeitung vor sehr große Herausforderungen. Zum einen ist es technisch sehr anspruchsvoll, große, sich schnell ändernde, und häufig auch sehr inhomogene Daten zu verarbeiten. Dies fasst man häufig zusammen unter den "vier V's der Big Data" *Volume, Velocity, Variety, and Veracity*. Zum anderen ist auch die Interpretation der Daten oft schwierig und erfordert komplexe Modelle. Das angestrebte Ideal ist deshalb eine fast interaktive Auswertung und Exploration von komplexen und sehr großen Daten, um so neue Erkenntnisse zu gewinnen. Insbesondere müssen hier notwendigerweise der informatische und der mathematische Aspekt harmonisch zusammenspielen, um Zugriff und Analyse dieser großen Datenmengen zu ermöglichen. Daher ist eine enge Zusammenarbeit und Verzahnung mit dem Masterstudiengang "Mathematics in Data Science" der Fakultät für Mathematik vorgesehen.

Reagierend auf diese Entwicklung richtete deshalb die Fakultät für Informatik der TUM einen Masterstudiengang „Data Engineering and Analytics“ ein. Parallel dazu richtete die Fakultät für Mathematik den Masterstudiengang „Mathematics in Data Science“ ein. Die beiden Studiengänge behandeln aufbauend auf einer gemeinsamen Grundlage speziell die informatischen bzw. mathematischen Aspekte von Big Data. Die Sammlung, Modellierung, Speicherung, Verarbeitung und Auswertung extrem großer (Beispiel: Soziale Medien), sich schnell ändernder (Beispiel: Sensoren) und komplexer (Beispiel: ökologische Systeme) Datenmengen ist das Ziel beider Studiengänge.

Auf der Informatikseite wird mehr Wert darauf gelegt, die Daten so zur Verfügung zu stellen, dass

sie mit unterschiedlichen Methoden zu unterschiedlichen Zwecken effizient verarbeitet werden können. Die Datenspeicherung und der Datenzugriff müssen auf die gewaltige Menge, die hohe Änderungsrate und die große Komplexität der Daten abgestimmt sein und variabel reagieren können. Weiterhin müssen sie sich den wechselnden Analysealgorithmen sowie den Visualisierungen der Ergebnisse anpassen. Der Informatikstudiengang soll sowohl ein tiefgreifendes Verständnis der grundlegenden Methoden als auch praktische Techniken zur Verarbeitung sehr großer Datenmengen vermitteln. Im Gegensatz dazu wird auf der Mathematikseite im Studiengang "Mathematics in Data Science" der Fokus auf die Berechnung, Simulation und Vorhersage komplexer Phänomene (wie z.B. Kundenverhalten, Wirtschaftsentwicklung, medizinische Krankendaten,...) und auf die oft schwierige Interpretation dieser Daten gelegt, welche komplexe mathematische Modelle erfordert. Damit unterscheiden sich beide Studiengänge in wesentlichen Teilen.

„Big Data“ steht für die Möglichkeit, die Datenflut nahezu in Echtzeit zu analysieren. Aktuelle Beispiele zeigen die Notwendigkeit und Bedeutung von Datensparsamkeit und Datenschutz. Ergänzend zu den technischen Aspekten werden deshalb auch die damit einhergehenden individuellen und gesellschaftlichen Problemstellungen adressiert.

Ziel des Studiengangs „Data Engineering and Analytics“ ist es daher, die Absolventen zu den oben beschriebenen informatischen und gesellschaftlichen Anforderungen des praktischen Umgangs mit großen Datenmengen zu befähigen. Die Beschäftigung mit großen Datenmengen erfordert neue Techniken und Algorithmen, um die klassischen Fragestellungen wie Datenbankabfragen, Datenanalyse usw. in diesem neuen Umfeld effizient bearbeiten zu können. Diese neuen Techniken und Verfahren soll der Studiengang zur Verfügung stellen.

## 1.2 Strategische Bedeutung des Studiengangs

Big Data in der strategischen Ausrichtung von Lehrstühlen: Insbesondere der Lehrstuhl für Datenbanksysteme soll auf den Bereich Data Engineering ausgerichtet werden. Dies zeigt sich in der Neubesetzung der Professur "Data Science and Engineering" durch Professor Thomas Neumann.

Big Data in der strategischen Ausrichtung der Informatikfakultät: Der Studiengang fügt sich passgenau in die Lehrstrategie der Fakultät für Informatik ein. Darin wird festgelegt, dass die TUM-Informatik neben der breiten Vertretung der Kernthemen insbesondere Themen verfolgt mit hohem Nachhaltigkeitspotenzial (z.B. Bioinformatik), nach ihrer Bedeutung in anderen Disziplinen/Fachkulturen (z.B. Informatik: Games Engineering), oder spezifischen Berufsbildern (z.B. CSE), oder wenn aus dem Professorenkollegium erforderliche Lehrveranstaltung angeboten werden können; es wird darin aber auch betont, dass spezielle Studiengänge, die die oberen Kriterien nicht mehr im gewünschten Umfang erfüllen, eingestellt werden können. Neben Robotics, Biomedical Computing, Games Engineering und Wirtschaftsinformatik sieht die Fakultät für Informatik die Beschäftigung mit Big Data einerseits als notwendigen Baustein der Informatik selbst an, andererseits aber auch als Querschnittsthema, das fachübergreifende Bedeutung hat. So spielt die Entwicklung von Big Data-Methoden innerhalb der Informatik eine wichtige Rolle wie auch die Anwendung von Big Data-Verfahren in anderen Fachgebieten. Der Master vertieft die Big Data-Aspekte, die in den Studiengängen Informatik, Wirtschaftsinformatik, Bioinformatik, Robotics, Biomedical Computing und CSE in kleinerem Umfang oder als Wahlmöglichkeit bestehen, zu einem eigenständigen, erweiterten Lehrangebot, das nicht nur die fachspezifischen, sondern alle relevanten Aspekte und Entwicklungsmethoden behandelt. Dadurch wird zusätzlich das Querschnittsthema Big Data gestärkt, und das Angebot dieser schon vorhandenen Studiengänge wird erweitert. Unterstützt und gefördert wird der Studiengang durch die neue Professur *Data Science and Engineering*, welche maßgebliche Forschungs- und Lehrbeiträge im Bereich Big Data liefert. Zusätzlich hat im August 2015 ein neuer Professor (Prof.

Dr. Günnemann) seine Forschungsgruppe *Data Mining & Analytics* in der Fakultät für Informatik eingebracht.

Big Data in der strategischen Ausrichtung der TUM: Der Umgang mit großen Datenmengen ist ein wichtiger Teil der Initiative "Zentrum Digitalisierung.Bayern" der Staatsregierung mit Sitz in Garching. Die TUM ist hierbei mit einer führenden Rolle eingebunden.

Das Thema Big Data spielt in verschiedensten Anwendungen eine wichtige Rolle. So sollen in Zukunft TUM-weit auch andere Fakultäten mit eingebunden werden, die z.B. Module zum Thema Big Data anbieten oder abrufen. Das Vorlesungs- und Projektangebot ist bereits für Studenten anderer Fakultäten attraktiv und wird genutzt. Wir sind in der Lage auf Kooperationsanforderungen anderer Fakultäten einzugehen.

## 2 Qualifikationsprofil

Die Anforderungen an den Masterstudiengang Data Engineering and Analytics gliedern sich, angelehnt an HQR, in (i) Wissen und Verstehen, (ii) Einsatz, Anwendung und Erzeugung von Wissen (iii) Wissenschaftliches Selbstverständnis/Professionalität und (iv) Kommunikation und Kooperation.

Im Berufsleben, nach erfolgter Ausbildung als Data Engineer<sup>1</sup> and Analyst<sup>2</sup>, sind unsere Absolventen in der Lage, z.B. in Technologieunternehmen hoch-skalierbare Lösungen zu entwickeln und zu verwenden und dabei auf cloud-scale Daten auszuwerten und zu analysieren (i), Trends mithilfe von Data Mining zu identifizieren, komplexe Anfragen zu optimieren, und dabei sowohl klassische Techniken wie Hadoop als auch cutting-edge Technologie für Multi-Terabyte Datenmengen einzusetzen (ii). Nach erfolgreichem Studium besitzen unsere Absolventen das dafür benötigte Wissen und eine tiefe Einsicht (iii) in informatischen Methoden und Algorithmen zur Datenanalyse und –aufbereitung wie R, SAS Enterprise Miner oder SPSS Clementine sowie mit Werkzeugen wie Hadoop, Map/Reduce oder Hive, und zu Datensicherheit. Außerdem verfügen sie über Kenntnisse in mathematischen Methoden der Computational Statistics und des Data Mining. Dadurch sind Absolventen des Masterprogramms in der Lage, anspruchsvolle Big Data-Aufgaben aus Natur- oder Ingenieurwissenschaften, wie z.B. Klassifikation, Suche und Erkennen von Ausreißern in sehr großen Datenmengen durch den Einsatz moderner und effizienter Methoden aus der Informatik und Mathematik erfolgreich zu bearbeiten (iv). Darüber hinaus besitzen die Absolventen ein Bewusstsein für Datensparsamkeit und Datenschutz (iii).

Die Absolventen zeichnen sich besonders durch folgende Kompetenzen aus:

- **Entwicklung moderner hochskalierbarer Datenverarbeitungstechniken:** Die Absolventen kennen den aktuellen Stand der Technik bzgl. Datenhaltung und Verarbeitung großer Datenmengen. Sie können diese Techniken einordnen und auch im Zusammenhang mit

1 Data engineers are the designers, builders and managers of the information or "big data" infrastructure. They develop the architecture that helps analyze and process data in the way the organization needs it. And they make sure those systems are performing smoothly (Quelle: Data Science Central, the online resource for Big Data practitioners, <http://www.datasciencecentral.com/profiles/blogs/data-scientists-vs-data-engineers> )

2 The data analysts job description includes importing, cleaning, transforming, validating or modeling data with the purpose of understanding or making conclusions from the data for decision making purposes. It may include presenting data in charts, graphs, tables, designing and developing relational databases for collecting data (Quelle: Data analysts online training, <http://www.dataanalystbootcamp.com/> ) .

klassischen Datenbanktechniken anwenden. Darauf aufbauend entwickeln sie selbständig neue Datenverarbeitungstechniken insbesondere zur hochskalierbaren Datenverarbeitung in verteilten Umgebungen.

- **Verstehen von komplexen Analysetechniken:** Die Absolventen verstehen klassische mathematische Analyseverfahren und Klassifikationstechniken, z.B. Support Vector Machines und Decision Trees. Diese Techniken bilden die mathematischen Grundlagen für die Weiterverarbeitung.
- **Anwendung von Analysetechniken zur Klassifikation von komplexen Daten:** Mit Hilfe der skalierbaren Datenverarbeitungstechniken können unsere Absolventen die mathematischen Verfahren anwenden und damit große Datenmengen effizient analysieren. Dabei analysieren sie die Anwendung sowohl hinsichtlich Effizienz als auch der Güte des Ergebnisses und korrigieren dann auch geeignet die Auswertungsstrategie.
- **Bewertung der Qualität und Aussagekraft der Analyseergebnisse:** Die Absolventen überprüfen die gewonnenen Aussagen bzgl. Qualität und Aussagekraft für die jeweiligen Fragestellungen. Dabei wählen sie iterativ geeignete Verfahren und Analysetechniken.
- **Gesellschaftliche Teilhabe:** Die Absolventen kennen die gesellschaftliche Bedeutung von Datensparsamkeit und Datenschutz. Sie reflektieren über die Auswirkung ihrer Analysen und kennen geeignete Techniken um Datensicherheit und Datensparsamkeit herzustellen.
- **Erlernen vertiefter englischsprachiger Fachkenntnisse:** Die Absolventen verstehen die englische Fachsprache im Bereich der Datenanalyse und -verarbeitung. Durch die durchgängig englischsprachigen Veranstaltungen sind sie optimal auf ihre zukünftige Berufstätigkeit in internationalen Unternehmen vorbereitet.
- **Befähigung zur Forschung:** Zusammenfassend sind die Absolventen in der Lage, auf Basis des aktuellen Stands der Forschung, Fragestellungen bzgl. weiterführender Probleme zu identifizieren, Forschungshypothesen zu formulieren und einen Forschungsplan aufzustellen. Sie können ihre Forschung und die Resultate in angemessener Weise schriftlich und mündlich sowohl gegenüber der eigenen wissenschaftlichen Community als auch gegenüber Laien kommunizieren. Durch die Erlangung von Planungs-, Forschungs- und Kommunikationskompetenz sind die Absolventen in der Lage, fundierte und richtungsweisende Entscheidungen zu treffen.

Im Vergleich zu Absolventen des Mathematik-Studiengangs "Mathematics in Data Science", die sich mehr auf mathematische Modellierung und Entwicklung von Analyseverfahren spezialisieren, können die Absolventen des Informatik-Studiengangs effizient sehr große Datenmengen verwalten und Analysetools anwenden, und setzen dabei fortgeschrittene Techniken der Informatik ein. Im Gegensatz zu Absolventen aus einer bestimmten Anwendungsdisziplin, z.B. Biologen, können Absolventen des Studiengangs „Data Engineering and Analytics“ nicht nur die konkreten Anwendungsaufgaben, z.B. in der Biologie, analysieren, von ihnen abstrahieren und die angemessenen Techniken der Informatik zur Problemlösung in dem speziellen Anwendungsfall für die effiziente Verwaltung und Analyse der großen Datenmengen auswählen und implementieren; sondern sie können die anstehenden Aufgaben auch besser lösen, da sie zusätzlich über ein tieferes Verständnis der mathematischen und informatischen Grundlagen verfügen.

## 3 Zielgruppen

### 3.1 Adressatenkreis

Das Studienangebot zielt auf Bachelor Informatik - Absolventen und Bachelor Mathematik - Absolventen mit Nebenfach Informatik (oder Ähnlichem), die sich in Data Engineering and Analytics spezialisieren wollen. Neben den Kenntnissen der Programmierung und der methodischen Grundlagen der Informatik setzt der Masterstudiengang Kenntnisse in mathematischen Methoden voraus. Es werden gute Kenntnisse der englischen Sprache erwartet. Alle wichtige Fachliteratur ist in Englisch, die Module sind fast alle in Englisch, der Arbeitsmarkt für die Absolventen ist international mit der Arbeitssprache Englisch. Der Studiengang richtet sich sowohl an TUM-Absolventen als auch an solche außerhalb der TUM, insbesondere auch aus dem Ausland. In Anfragen sehen wir ein großes Interesse an einem Big Data Studiengang, insbesondere aus dem nicht-europäischen Ausland sowohl von Studierenden als auch von Partneruniversitäten. Internationale Studierende sollen ein Mindestmaß an Kenntnissen der deutschen Sprache mitbringen oder zeitnah zu Beginn ihres Studiums erwerben. Dies ist wichtig für die Integration in den Studienalltag, in dem die Studierenden mit deutschen Kommilitonen zusammen studieren, Arbeits- und Projektgruppen bilden, und sich zwar im professionellen Umfeld in der Fachsprache Englisch verständigen, aber im alltäglichen Umgang auch immer mal wieder auf Deutsch gewechselt wird. Auch in alltäglichen Situationen im privaten oder beruflichen Umfeld, z.B. als Werkstudierende, werden sie mit Deutsch konfrontiert und müssen rasch lernen, sich hier zurechtzufinden.

### 3.2 Vorkenntnisse der Studienbewerber

Vorausgesetzt werden elementare Kenntnisse aus der Informatik, insbesondere methodische Grundlagen der Informatik, des Programmierens, von Algorithmen und Datenbanken. Weiterhin werden elementare Kenntnisse in mathematischen Methoden erwartet, entsprechend der TUM Module Diskrete Strukturen, Diskrete Wahrscheinlichkeitstheorie, Analysis und Lineare Algebra. Da der Bewerberkreis auch Absolventen anderer Abschlüsse als Informatik und auch ausländische Studierende umfassen soll, und daher sehr heterogen zusammengesetzt sein wird, soll in einem Eignungsverfahren sichergestellt werden, dass die Kompetenzen der Bewerber trotzdem den speziellen Anforderungen des Studienganges genügen.

### 3.3 Zielzahlen

Das Ziel besteht darin, allen geeigneten Bewerbern einen Studienplatz anzubieten. Zum Wintersemester 2017 wurden 96 Studierende zugelassen. Davon haben 68 den Studienplatz angenommen.

## 4 Bedarfsanalyse

Die ohne jeden Zweifel vorhandenen großen, wichtigen, zu erforschenden Datenmengen haben einen unschätzbaren wirtschaftlichen Wert, z.B. wird Big Data als das „Öl der Zukunft“ bezeichnet, und führen zu breiten gesellschaftspolitischen Diskussionen, z.B. hinsichtlich der Macht der sozialen Netzwerke oder der Geheimdienste.

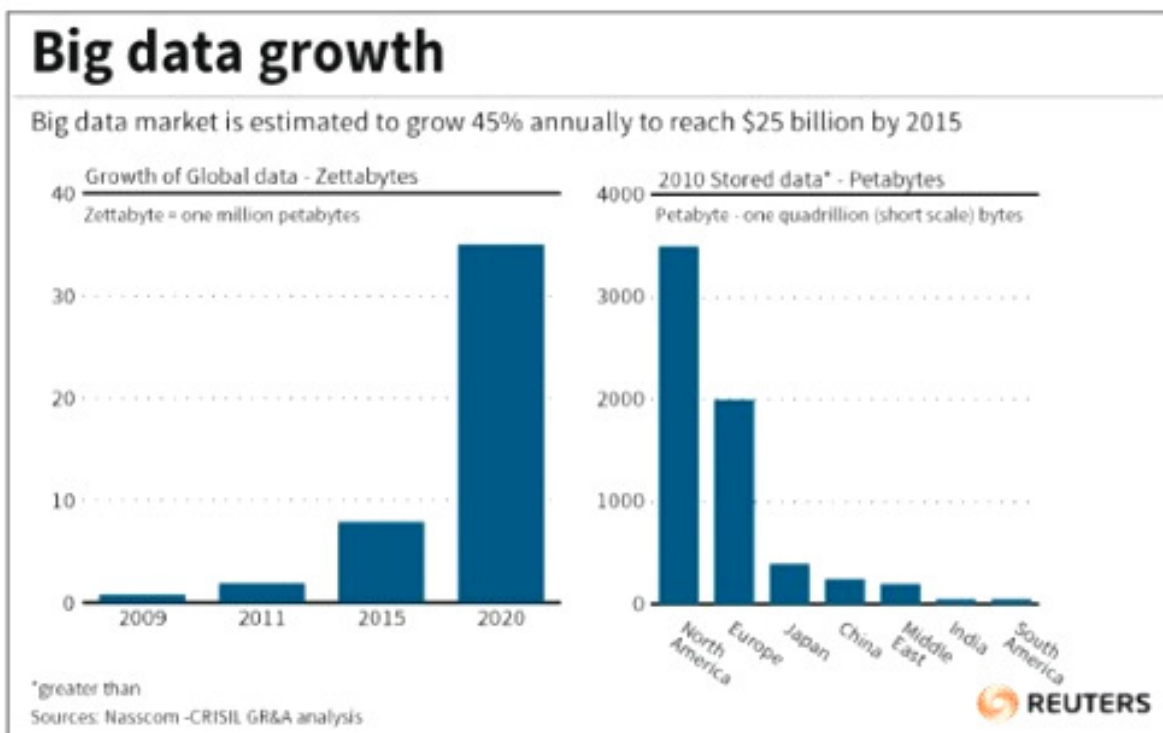
Im Oktober 2012 lautete der Titel des Harvard Business Review: Data Scientist: The Sexiest Job of the 21st Century. Auszug: "...Data scientists are the key to realizing the opportunities presented by big data. They bring structure to it, find compelling patterns in it, and advise executives on the implications for products, processes, and decisions. They find the story buried in the data and communicate it. And they don't just deliver reports: They get at the questions at the heart of problems and devise creative approaches to them...". In der Studie "Worldwide Big Data Technology and Services, 2012–2015 Forecast" durchgeführt vom IDC (International Data

Corporation), ist für Big Data Technologien und Anwendungen weltweit mit einer jährlichen Wachstumsrate von 40% zu rechnen – ungefähr das Siebenfache wie für den ICT (Information and Communication Technology) Markt insgesamt.

Im Berufsleben nach erfolgter Ausbildung als *Data Engineer and Analyst* werden unsere Absolventen z.B. im Bereich Marketing, Logistics und Customer Service arbeiten, um das Kundenverhalten zu analysieren, Trends herauszufinden mittels Data Mining, SQL-Anfragen zu optimieren basierend auf cutting-edge Technologie und Multi-Terabyte Datenmengen. Aus diesem Grund sind Firmen wie *Allianz, Siemens* (zu beiden bestehen schon Kontakte im Bereich Data Science), *Munich Re, check24, immobilienscout24, Zalando* sehr stark daran interessiert, entsprechende Absolventen anzustellen, um z.B. aus den Kundendaten maßgeschneiderte Angebote/Vorhersagen liefern zu können und sich möglichst genau den Kundenwünschen anpassen zu können. Besonders im Münchner Umfeld sind hier viele Firmen tätig, die entsprechende Absolventen übernehmen werden.

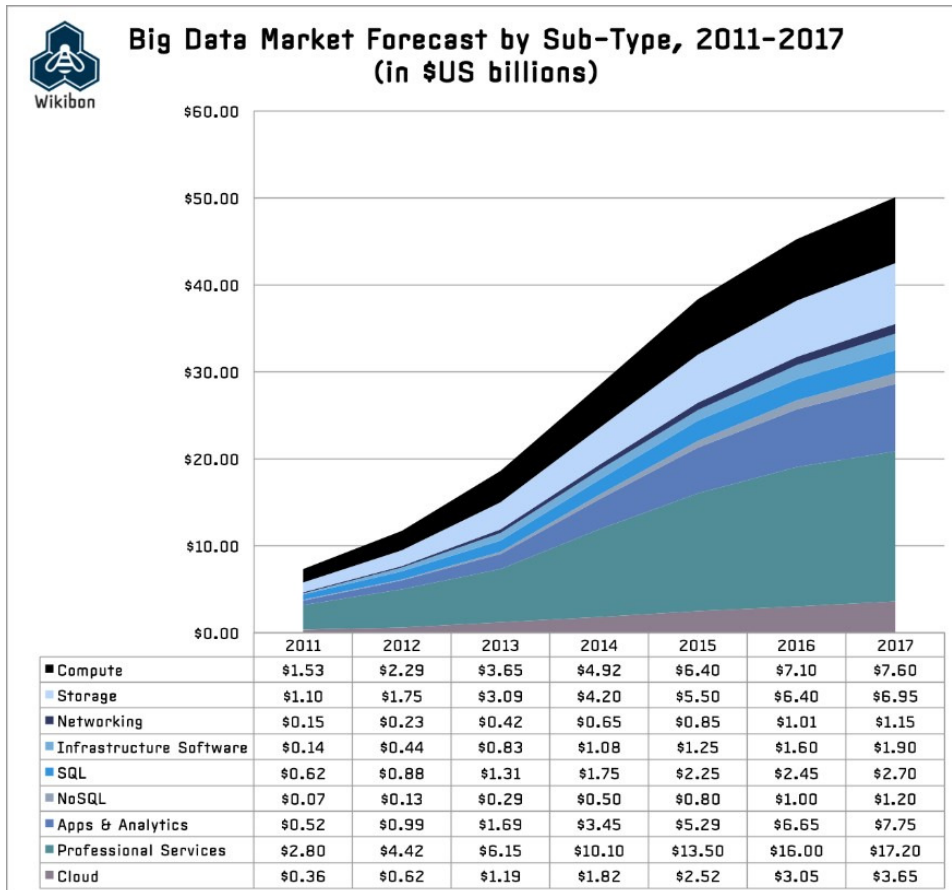
Weiterhin muss sich hier die TUM den relevanten Firmen, die ja vor allem in München beheimatet sind, als kompetenter Ansprechpartner präsentieren, der dieses aktuelle Thema wissenschaftlich und lehrmäßig abdeckt.

Das Anwachsen der Nachfrage an Big Data-Spezialisten auf dem Arbeitsmarkt ist ersichtlich aus der folgenden Graphik, die das Wachstum des entsprechenden Marktes aufzeigt.



Reuters.graphic/Catherine Trevethan 05/10/12





## 5 Wettbewerbsanalyse

Im Gegensatz zu den USA, wo Präsident Obama schon im März 2012 die „Big Data Initiative“ ausrief und es inzwischen schon einige Masterprogramme in Data Science gibt, kommt in Deutschland bisher erst langsam Bewegung in die Ausbildungslandschaft. Im Juli 2014 startete die Europäische Kommission eine „New strategy on Big Data“. In Deutschland wurde die „Big Data Allianz“ der Fraunhofer Institute ins Leben gerufen. Die DFG hat neue Programme aufgelegt wie „Algorithms for Big Data“ SPP 1736 und „Compressed Sensing in der Informationsverarbeitung“ SPP 1798.

Studiengänge im Umfeld des Big Data werden zurzeit vermehrt angeboten, bzw. neu geschaffen. In Deutschland, insbesondere im Raum München, besteht hier jedoch großer Bedarf.

### 5.1 Externe Wettbewerbsanalyse

In den USA z.B. gibt es eine ganze Reihe von Universitäten, die Studienprogramme im Bereich Data Science anbieten, ebenso in Australien, wenn auch erheblich weniger, in Südkorea gibt es auch ein Masterprogramm, Singapur baut auf die Data Science Spezialisierung von der Johns Hopkins University. In Europa werden z.B. an der ETH Zürich Professuren im Bereich Big Data besetzt, aber es gibt keinen Studiengang hierzu. Programme, die ähnlich dem Studiengang Data Engineering and Analytics sind, gibt es z.B. in Dänemark; hier werden jedoch die Mathematikgesichtspunkte vernachlässigt. Am nächsten an den Masterstudiengang kommt noch der Eurotech-Partner TU Eindhoven mit seinem Studienprogramm Data Science in Engineering, hier wird auch explizit die Verbindung der Informatik mit der Mathematik betont. Auch das Weiterbildungsprogramm der Ecole Polytechnique, Data Science and Big Data, verbindet Informatik und Mathematik.

Im internationalen Vergleich sieht man deutlich, dass Masterstudiengänge im Bereich Big Data im Vormarsch sind. Auch unsere Partneruniversitäten haben schon einschlägige Studiengänge etabliert oder überlegen, dies zu tun. In allen Gesprächen der letzten Zeit mit Partneruniversitäten war die Ausbildung im Bereich Big Data ein Thema. Hier musste die TUM aufschließen und der Studiengang Data Engineering and Analytics schließt einerseits die Lücke, setzt aber auf der anderen Seite auch durch die Verbindung von Informatik und Mathematik eigene wichtige Akzente.

In Deutschland sind die einschlägigen Studienangebote sehr begrenzt. Die wichtigsten sind der Studiengang Datenwissenschaften in Dortmund, welcher in der Fakultät für Statistik angesiedelt ist, Data Engineering an der Jacobs University in Bremen und Data Science an der Universität Potsdam. Bisherige Studienangebote beschränken sich meist auf einen der beiden Aspekte Informatik oder Mathematik. Im Bereich des Data Engineering hingegen gibt es nur sehr wenige andere Studienprogramme; das Programm der Jacobs University Bremen ist eine Ausnahme, es ist jedoch sehr klein (in allen vier Masterprogrammen der Universität sind ca. 150 Studierende eingeschrieben) und es ist gebührenpflichtig. In der Ausbildung in deutschen Informatikstudiengängen wird zwar in den unterschiedlichen Bereichen auf die Herausforderung des Data Engineering eingegangen. Es hat sich aber noch kein kanonischer Studiengang entwickelt.

Auch die LMU hat einen neuen Studiengang im Bereich Data Science in der Fakultät für Mathematik, Informatik und Statistik eingerichtet. Es werden hier sowohl Statistik- als auch Informatikmodule vorhanden sein. Die Schwerpunkte in der Informatik an der LMU haben gegenüber denen der TUM eindeutig eine stärkere mathematische Ausrichtung. Hingegen wird bei unserem Studiengang sich davon absetzend die ingenieurmäßige Anwendung von Big Data im Fokus stehen. Insbesondere die effiziente Speicherung und Verarbeitung sehr großer, oft verteilter Datenmengen ist in der Praxis eine große Herausforderung, bedingt die Erstellung eines komplexen Softwaresystems und wird deshalb in dem Studiengang umfassend behandelt. Zusätzlich bietet die enge Zusammenarbeit mit dem Mathematik-Studiengang "Mathematics in Data Science" an der TUM ein breiteres, umfassenderes Angebot an Big Data-relevanten Themen.

## 5.2 Interne Wettbewerbsanalyse

Der Informatik-Studiengang unterscheidet sich von schon bestehenden nationalen und internationalen Programmen durch das ganzheitliche Angebot mit dem Hauptschwerpunkt auf informatischen Aspekten (insbesondere sowohl Datenspeicherung und –zugriff), aber auch einem erheblichen Anteil an mathematischen Aspekten (insbesondere Datenanalyse und –auswertung). Beim Mathematik-Studiengang sind die Schwerpunkte genau anders verteilt, aber auch beide vorhanden. Im Anhang werden in einer Strukturübersicht beide Studiengänge mit ihren Gemeinsamkeiten und Unterschieden grafisch dargestellt. Der Informatik-Studiengang vermittelt Lösungen, die Daten so zur Verfügung zu stellen, dass sie mit unterschiedlichen Methoden zu unterschiedlichen Zwecken effizient verarbeitet werden können. Die Speicherung und der Zugriff müssen auf die gewaltige Menge, die hohe Änderungsrate und die große Komplexität der Daten abgestimmt sein und variabel reagieren können. Weiterhin müssen sie sich den wechselnden Analysealgorithmen sowie den Visualisierungen der Ergebnisse anpassen. Der Mathematik-Studiengang beschäftigt sich vertieft mit mathematischen Analysemethoden und Fragestellungen, z.B. zur Regressionsanalyse für den statistischen Zusammenhang zwischen unterschiedlichen Attributen oder der Visualisierung von Analyseergebnissen.

Auf der anderen Seite garantiert der Mathematikanteil, zusammen mit den neu angebotenen Labs und Projekten, dass der Studiengang „Data Engineering und Analytics“ sich grundlegend von einem reinen Informatik-Master unterscheidet. Im Gegensatz zum klassischen Informatik-Master liegt der Schwerpunkt auch auf dem Umgang mit extrem großen Datenmengen, was eine

spezielle Herausforderung darstellt und besondere Kenntnisse z.B. von entsprechenden Datenbanken, Verteilten Daten und probabilistischen Methoden erfordert. In der Wirtschaftsinformatik wird das Thema Big Data auch schon aufgegriffen. So gibt es im Masterstudiengang Wirtschaftsinformatik den Schwerpunkt Business Analytics and Optimization. In Abgrenzung hierzu soll der neue Studiengang den Schwerpunkt auf den reinen Informatikaspekt des Data Engineering legen, und sich erst darauf aufbauend auch mit Business Analytics-Gesichtspunkten beschäftigen. Der Master Computational Science and Engineering (CSE) hat zwar als Anwendungskatalog auch Big Data, der Studiengang richtet sich aber gerade nicht an Informatik- oder Mathematik-Absolventen, sondern an Studierenden, die einen Abschluss in einem Anwendungsfach haben. Der Masterstudiengang Bioinformatik beschäftigt sich mit einem speziellen Aspekt von Big Data, nämlich der Anwendung in der Biologie, insbesondere z.B. der Genom-Sequenzanalyse, während in "Data Engineering and Analytics" allgemeine Methoden und deren Entwicklung im Vordergrund stehen.

## 6 Aufbau des Studiengangs

Der Studiengang umfasst als Vollzeitstudiengang (Teilzeitvariante ev. später) vier Semester (120 Credits). Der Studiengang startete erstmals zum Wintersemester 2016/2017, seitdem ist der Studienbeginn sowohl im Winter- als auch im Sommersemester möglich. Der Aufbau des Studiengangs richtet sich an seinem Qualifikationsprofil aus. Die Ausbildung hat das Ziel, die Studierenden auf der Basis vermittelter Methoden und Kompetenzen zu eigenständiger Forschungsarbeit anzuregen. Aufgrund der Tatsache, dass sämtliche Pflichtmodule nur in englischer Sprache angeboten werden und die Abschlussarbeit in englischer Sprache zu verfassen ist, verfügen die Absolventen über vertiefte englischsprachige Fachkenntnisse sowohl in Wort als auch in Schrift. Als besonders wichtig erscheint es auch, die Studierenden auf die Möglichkeiten vorzubereiten, selbstverantwortlich tätig zu werden, z.B. durch Umsetzen neuer Ideen in eigenen Analyseverfahren. Aber auch Firmenkontakte und praktische Anwendungen sollen gepflegt werden, insbesondere aus dem breiten Münchner Stellenmarkt, welcher den Studierenden beste Möglichkeiten bietet, schon während des Studiums wichtige einschlägige praktische Erfahrungen zu sammeln und diese mit exzellenter Forschung zu verknüpfen. Auf der Basis **vertieften Grundlagenwissens** konzentrieren sich die Studienziele vor allem auf ein **an den aktuellen Forschungsfragen orientiertes Fachwissen** und die Erlangung (berufsrelevanter) Schlüsselqualifikationen; ein Schwerpunkt ist dabei die Entwicklung moderner hochskalierbarer Datenverarbeitungstechniken. Die Module des Studiengangs beziehen in den jeweiligen Fachgebieten aktuelle Forschungsergebnisse und Forschungsrichtungen ein, wodurch die Studierenden Planungs- und Forschungskompetenzen erhalten. So werden z.B. aktuelle Entwicklungen in der Prozessorarchitektur verbunden mit Fragestellungen im Bereich Big Data, u.a. in der hier entscheidenden skalierbaren Anfrageoptimierung. Die Einbeziehung unterschiedlicher Informatik-Fachgebiete im Hinblick auf die Bewältigung der Probleme im Bereich Big Data erfordert einen tiefen Einblick in die einzelnen Gebiete und die Fähigkeit, diese planvoll und zielgerichtet zu verbinden sowie hierdurch neue Kenntnisse zu erlangen. Ein Forschungsprojekt vermittelt die Vertiefung der neuen Kenntnisse anhand eines komplexen Forschungsproblems im Bereich Big Data. Durch die Ablegung eines Anwendungsprojektes andererseits wird die Übertragung dieser Fähigkeit in ein anderes Fachgebiet geschult und Kommunikationsmaßnahmen vermittelt, so dass die Absolventen eine Brücke in Anwendungsbereiche von Big Data schlagen können. Ein Forschungsprojekt oder ein Anwendungsprojekt ist wahlweise abzulegen.

Der Studiengang teilt sich grob in drei Bereiche:

- Pflichtbereich (A)
- Wahlbereich (B)
- Unterstützende Wahlmodule (D).

Im **Pflichtbereich** gibt es je ein grundlegendes Pflichtmodul für den Bereich Data Engineering und den Bereich Data Analysis. Beide sind Pflichtmodule in den beiden Studiengängen *Data Engineering und Analytics* und *Mathematics in Data Science*; ersteres wird von der Fakultät für Informatik, zweiteres von der für Mathematik angeboten. Nach der Ablegung der beiden Module verstehen die Studierenden die grundlegenden Techniken von Datenverarbeitung und –analyse sowohl im Bereich Informatik als auch Mathematik; sie sind in der Lage diese konkret auf Problemstellungen anzuwenden.

Zusätzlich gibt es – genau wie auch im Studiengang *Mathematics in Data Science* – ein Seminar und ein Praktikum.

Die Module des **Wahlbereichs** lassen sich in drei unterschiedliche Kategorien von Big Data einordnen:

- Data Engineering (B1)
- Data Analytics (B2)
- Data Analysis (B3)

Alle diese drei Kategorien sind wichtige Bestandteile einer gründlichen Ausbildung im Bereich Big Data, so dass die Studierenden zusammen mit den Pflichtmodulen in allen drei Kategorien vertieftes Grundlagenwissen und Kompetenzen erwerben. Die drei Kategorien werden ihrerseits wiederum in einen **grundlegenden** Teil und in einen **fortgeschrittenen** Teil aufgeteilt, wobei der vertiefende Teil von Data Analysis dem Studiengang *Mathematics in Data Science* vorbehalten ist – umgekehrt ist der fortgeschrittene Teil von Data Engineering dem Studiengang *Data Engineering and Analytics* vorbehalten. Wesentliche Ziele des grundlegenden Teils des Wahlbereichs sind insbesondere ein tiefes Verständnis und eine eigenständige Weiterentwicklung von hochskalierbarer verteilter Datenverarbeitung, z.B. aus dem Bereich Data Engineering durch die Module Query Optimization oder Distributed Systems. Die Bewertung der Qualität und Aussagekraft von Analyseergebnissen wird erreicht durch die Module im Bereich Data Analytics, z.B. Machine Learning oder Business Analytics. Im Bereich Data Analysis werden Module aus dem Studiengang *Mathematics in Data Science* eingebunden, wodurch sichergestellt wird, dass die erforderlichen Grundkenntnisse in mathematischer Datenanalyse erworben werden.

Der Pflicht- und der grundlegende Teil des Wahlbereichs stellen sicher, dass die wesentlichen Qualifikationsziele des Studiengangs erreicht werden. Eine Vertiefung nach individuellen Interessen ist im **fortgeschrittenen Teil der beiden Bereiche** Data Engineering und Data Analytics möglich. Eine Spezialisierung auf den Bereich Data Analysis ist nur im Studiengang *Mathematics in Data Science* möglich. Im weiterführenden Bereich wird hauptsächlich an aktuellen Forschungsfragen orientiertes Fachwissen vermittelt.

Die Studierenden identifizieren für Big Data relevante Forschungsfragestellungen und entwickeln eigene Lösungsansätze, z.B. im Modul Advanced Computer Networking aus dem Bereich Data Engineering. In diesem Bereich werden auch zwingend interdisziplinäre Kompetenzen gebildet. Die Studierenden arbeiten an der Schnittstelle zur Anwendung an einem Projekt, welches entweder eine Forschungsarbeit ist oder sie wenden Analysetechniken zur Klassifikation komplexer Daten auf konkrete Problemstellungen außerhalb der Kern-Informatik an. Sie werden so auf eine spätere Berufstätigkeit in angrenzenden Bereichen vorbereitet. Problemstellungen außerhalb der Informatik ermöglichen neuartige Daten und Anwendungsbereiche, die über klassische Fragestellungen hinausgehen, z.B. Analyse von seismologischen Daten zur Lokalisierung von Erdbeben.

Im Bereich Data Analytics sind neben vertiefenden Modulen wie z.B. Parallele Algorithmen auch relevante Module aus anderen Fakultäten, wie z.B. Information Retrieval in High Dimensional Data von der Fakultät für Elektrotechnik und Informationstechnik zu finden.

In den unterstützenden Wahlmodulen wird insbesondere auf eine Sensibilisierung im Hinblick auf den verantwortlichen Umgang mit großen Datenmengen hingewirkt. Die Studierenden besuchen hier mindestens einen der Kurse des MCTS im Bereich Social and Political Aspects of Data Science. Zudem kann in diesem Bereich z.B. die Fachsprachkompetenz gestärkt werden.

## 6.1 Pflichtbereich (A)

Die beiden Module IN2326 Foundations in Data Engineering (8CP), und MA4800 Foundation of Data Analysis (8CP), von denen das erste immer im Wintersemester von der Informatik und das zweite immer im Sommersemester von der Mathematik angeboten wird, bilden die spezifischen methodischen Grundlagen der beiden Studiengänge *Data Engineering and Analytics* und *Mathematics in Data Science*. Sie sind verpflichtend für alle Studierenden beider Studiengänge, da sie einen Überblick bieten und die Grundlagen in den beiden Eckpfeilern für Big Data, Data Engineering und Data Analysis legen. Die Absolventen erhalten mit diesen Grundlagen die Kompetenzen, die grundlegenden Konzepte der beiden Bereiche anzuwenden.

Sie befähigen die Studierenden Module im Wahlbereich aller drei Kategorien „Data Engineering“, „Data Analytics“ und „Data Analysis“ zu besuchen, auch wenn sie im Vorstudium bisher nur die in den Qualifikationsvoraussetzungen festgelegten Minimalvoraussetzungen in der Informatik bzw. der Mathematik erworben haben.

Durch das verpflichtende Seminar wird die Fähigkeit zur eigenständigen wissenschaftlichen Arbeit gefördert. Die praktische Umsetzung der erlernten Konzepte und Methoden wird im Rahmen eines verpflichtenden Praktikums vertieft. Ein neu geschaffenes Praktikum im Datenbankbereich ist der ideale Kandidat für die praktische Einübung der erlernten fortgeschrittenen Kenntnisse und Methoden. Weitere Master-Praktika aus der Informatik oder der Mathematik sind auf Antrag wählbar.

**Grundlagen - Pflichtbereich, 31 Credits:** Beide grundlegenden Module mit je 8 CP sowie ein Seminar (5 CP) und ein Praktikum (10 CP) sind Pflicht.

## 6.2 Wahlbereich (B)

### a) Grundlegend (B1.1, B2.1, B3)

Im **grundlegenden** Wahlbereich werden fortgeschrittene Kenntnisse erworben, die für die effiziente Verarbeitung von großen Datenmengen auf hohem methodischem Niveau erforderlich sind. Hierdurch sollen die Kompetenzen vermittelt werden, in unterschiedlichen Fachgebieten wissenschaftlich fundierte Entscheidungen zu treffen, zukünftige Forschungs- und Anwendungsfelder zu erkennen und bereichsspezifische Teilkomponenten einer komplexen Datenexploration eigenständig zu entwickeln, etwa die effiziente Ablage der Daten, die verteilte Datenhaltung, das systematische Durchforsten der Daten mit modernsten Informatikmethoden und komplexen mathematischen Modellen oder moderne Anwendungen wie Künstliche Neuronale Netze an der Schnittstelle zwischen Statistik und Informatik. Die Studierenden erlangen eine spezielle Fachexpertise, die sie im Vergleich zu den Absolventen des Masterstudiengangs Informatik in den entsprechenden Teilgebieten des Data Engineering und Analytics durch die entsprechende Fokussierung mit weit überdurchschnittlichen Kompetenzen ausstattet. Die Masterstudiengänge Informatik und Mathematik zielen auch auf eine breite Abdeckung des jeweiligen gesamten Fachgebietes ab. In der Informatik wird dies z.B. dadurch erreicht, dass aus mindestens drei unterschiedlichen Fachgebieten Module gewählt werden

müssen. Im Gegensatz hierzu gibt es im Studiengang *Data Engineering and Analytics* eine viel stärkere Fokussierung auf die Module des Teilbereichs der Informatik, der sich der Lösung von Problemen im Bereich Big Data widmet. Durch die weitgehend freie Wahl der einzelnen Module in diesem Teilbereich werden andererseits besondere Stärken und Interessen der Studierenden abgebildet und somit das Ausbildungsniveau durch ein eigenes Profil abgerundet.

Fortgeschrittene Konzepte verteilter Datenbanken (IN2140) und die Optimierung von Anfrageverarbeitung (IN2219) sind Module aus der Kategorie „**Data Engineering**“ (B1.1), die weiterführende angewandte Methoden behandeln und im sich weiter rasch entwickelnden Datenbankbereich wichtige langlebige Kompetenzen vermitteln. Eine weitere Kernkompetenz kann im Bereich verteilte Systeme (IN2259) erworben werden; insbesondere moderne Rechner- und Netzarchitekturen spielen eine entscheidende Rolle im Bereich Data Engineering. Hier ist auch eine Profilierung auf HPC (High Performance Computing, IN2013) möglich. Weitere Module tragen zu dem oben erwähnten Bereich bei.

Saubere Methodik in unterschiedlichen Dimensionen für die Exploration und Auswertung von komplexen Datenmengen kann in der Kategorie „**Data Analytics**“ (B2.1) in den Bereichen Mining Massive Datasets (IN2323), Maschinelles Lernen (IN2064) und Business Analytics (IN2028) erworben werden. Künstliche Intelligenz (IN2062), Modellbildung (IN2010) und Visualisierung (IN2026) ermöglichen die Exploration und Darstellung daraus resultierender Ergebnisse. Auch hier runden weitere Module den oben erwähnten Bereich ab.

In der Kategorie „**Data Analysis**“ (B3) werden wichtige mathematische Grundlagen für die Datenanalyse gelehrt. Sowohl die konvexe Optimierung (MA2504) als auch die Angewandte Regressionsanalyse (MA4401) bieten Methoden zur Analyse von großen Datenmengen. Weitere Module aus der Mathematik runden diesen Bereich ab. Hierbei ist von den Studierenden darauf zu achten, dass sie die notwendigen Vorkenntnisse, die in den Modulbeschreibungen festgehalten sind, mitbringen; einige Module haben Voraussetzungen, die über die minimal geforderten aus der Mathematik hinausgehen.

Durch die Festlegung einer Mindestcreditanzahl, die im Wahlbereich zu erbringen ist, wird die Fokussierung auf den Teilbereich der Informatik, der sich der Lösung von Problemen im Bereich Big Data widmet, erreicht. Durch die weitere Festlegung, dass alle drei Kategorien gewählt werden müssen, ist sichergestellt, dass keine Kategorie ganz abgewählt werden kann; Grundkenntnis in allen Kategorien ist für ein Kompetenzprofil im Bereich Big Data unerlässlich.

Die genaue Liste der Module und ihrer Zuordnung finden sich in der Anlage 1 der FPSO. Weitere Angebote zu aktuellen Themen werden lokal von der Studienkommission hinzugefügt.

**Grundlegender Wahlbereich (B1.1, B2.1, B3), mindestens 15 Credits:** mindestens ein Wahlmodul aus jedem der drei Kategorien „Data Engineering“, „Data Analytics“ und „Data Analysis“ muss gewählt werden.

b) Fortgeschritten (B1.2, B2.2)

Weiterführende Module können entsprechend den Interessen der Studierenden aus dem Wahlkatalog des fortgeschrittenen Bereichs Kategorie „Data Engineering“ (B1.2) und „Data Analytics“ (B2.2) gewählt werden.

In der Kategorie Data Engineering muss von den Studierenden eine Arbeit im Umfang von 10 CP gewählt werden. Zur Verstärkung der Forschungskompetenz können die Studierenden die essentiellen Aspekte eines wissenschaftlichen Teilbereichs von Data Engineering unter Anleitung analysieren und bewerten und sich eigene (eng begrenzte) Ergebnisse mit wissenschaftlichen Methoden erarbeiten (IN2169); dies befähigt die Studierenden noch intensiver für eine spätere

Forschungstätigkeit. Oder die Studierenden lernen exemplarisch einen Anwendungsbereich kennen. Da die Exploration und Auswertung komplexer Daten im späteren Beruf in vielen unterschiedlichen Anwendungskontexten benötigt wird, ist ein Projekt im Anwendungsbereich eine ergänzende Berufsbefähigung. Ein möglicher solcher Anwendungsbereich ist die **Wirtschaftsinformatik**. So ergänzen heutzutage (wenn nicht sogar ablösen) datengetriebene Geschäftsmodelle traditionelle Wertschöpfungsmodelle. Somit sind in diesem Bereich verstärkt Methoden des Data Engineering einzusetzen und ermöglichen somit einen entscheidenden Wettbewerbsvorteil. Ein weiteres Beispiel kommt aus der **Bioinformatik** bzw. der **Medizin**. Funktionen und insbesondere Wechselwirkungen von Genen müssen für neue Therapieansätze in der Medizin erforscht werden; hier sind enorme Datenmengen zu verarbeiten und intelligente Analysemethoden einzusetzen. Das Modul IN2328 (10 CP) beschreibt diese Projektarbeit.

In der Kategorie „Data Analytics“ können z.B. mathematische Optimierungsaspekte bei der Analyse von großen Datenmengen sowohl durch Module der Fakultät für Mathematik (z.B. MA4064 Fourier Analysis, MA4502 Kombinatorische Optimierung) als auch der Fakultät für Elektrotechnik (EI7223 Information Retrieval in High Dimensional Data) ergänzend hinzugenommen werden. Es ist hier aber auch möglich sich z.B. im Bereich Scientific Computing (durch IN2001) zu vertiefen.

Durch die Festlegung einer Mindestcreditanzahl, die im fortgeschrittenen Bereich zu erbringen ist, wird eine Vertiefung in den beiden Kategorien „Data Engineering“ und/oder „Data Analytics“ erreicht. Entweder ein Forschungs- oder ein Anwendungsprojekt ist zu absolvieren. Zudem sind mindestens weitere 15 Credits zu erbringen, was zur Profilierung in speziellen Aspekten von Big Data führt.

**Fortgeschrittener Wahlbereich (B1.2, B2.2), mindestens 25 Credits:** 10 CP für Forschungsprojekt IN2169 oder Anwendungsprojekt IN2328 (beide Module werden in vielen Ausprägungen zur Auswahl angeboten), alle weiteren Module sind Wahlmodule

Neben den Modulen des Wahlbereichs (grundlegend und fortgeschritten) sind auch alle Module aus dem Wahlkatalog des Masterstudiengangs Informatik (11 Fachgebiete insgesamt, <http://www.in.tum.de/en/current-students/masters-programs/informatics/elective-modules/fps0-2018/>) wählbar. Hierdurch kann das Profil um weitere Informatikbereiche angereichert werden.

**Weitere Module, mindestens 13 Credits:** aus dem Wahlbereich oder dem Wahlkatalog des Masterstudiengangs Informatik

### 6.3 Unterstützende Wahlmodule (D)

#### a) Überfachliche Grundlagen (D1)

Absolventen des Masterstudiengangs „Data Engineering and Analytics“ müssen verantwortungsvoll mit den Daten umgehen, die sie in ihrer Tätigkeit extensiv speichern und verarbeiten werden. Darauf wird zum einen in den jeweiligen Fachmodulen, insbesondere in dem Grundlagenmodul Foundations in Data Engineering eingegangen. Weiterhin haben die Studierenden die Möglichkeit, spezielle, zu Data Science passende Module aus dem Katalog der überfachlichen Grundlagen zu belegen. Gerade im Umgang mit großen Datenmengen, die verknüpft und exploriert werden, ist eine Sensibilisierung im Hinblick auf den verantwortlichen Umgang mit diesen Daten von besonderer Wichtigkeit. IN9044 Seminar Datenschutz eignet sich hierfür in besonderer Art und Weise. Aber auch Module der Carl von Linde Akademie, wie z.B. CLA30127 oder das Seminar CLA31212 How Do We See Big Data? können hierzu beitragen. Weiterhin können die Studierenden aber auch z.B. ihre Fachsprachenkompetenz in diesem

Bereich vertiefen. Ein Wahlkatalog, aus dem die Studierenden mindestens 3 Credits auswählen, wird vom Prüfungsausschuss zur Verfügung gestellt.

- b) Social and Political Aspects of Data Science (D2)

Das MCTS bietet spezielle Module zu Social and Political Aspects of Data Science an. Diese Module sind auch im Studiengang „Mathematics in Data Science“ verankert.

**Unterstützende Wahlmodule (D), mindestens 6 Credits:** je 3 CP aus dem Wahlkatalog von D1 und D2

## 6.4 Master's Thesis (C)

Die abschließende Masterarbeit (30 Credits) wird als Nachweis des wissenschaftlichen Handwerks verstanden und soll die Befähigung der Studierenden zeigen, neues Wissen zu produzieren. Diesbezüglich sollte die Masterarbeit ein integraler Bestandteil von laufenden Forschungsprojekten in einer der gewählten Kategorien sein. Die Studierenden schaffen eigene Resultate und können sie in angemessener Weise sowohl schriftlich als auch mündlich sowohl gegenüber der eigenen wissenschaftlichen Community als auch gegenüber Laien kommunizieren.

**Master's Thesis: 30 Credits**

## 6.5 Bereitstellung eines Mobilitätsfensters

Der Studienplan ist so strukturiert, dass er eine individuelle Gestaltung je nach persönlicher Schwerpunktsetzung erlaubt und somit auch ein Fenster für Auslandsaufenthalte bietet. Insbesondere eignet sich das 3. Semester als Mobilitätsfenster an, in dem die Studierenden ein Semester im Ausland verbringen können, ohne dass sich die Regelstudienzeit verlängert.

Die Fakultät für Informatik möchte ihren Studierenden nach Möglichkeit die Gelegenheit eines Auslandsstudiums bieten. Sie unterhält vielfältige Kontakte zu Partneruniversitäten im europäischen Ausland über Erasmus-Verträge und nimmt am weltweiten TUMexchange Programm teil. Eine vereinfachte Einbringung aus einem mit der TUM organisierten Auslandsaufenthalt während des Studiums wird im Umfang von bis zu 30 ECTS unterstützt. Zusätzlich zur üblichen Anerkennung von gleichwertigen Modulen bedeutet hier vereinfacht, dass es keine gleichwertigen Module an der TUM im Curriculum hierfür geben muss, sondern dass nur fortgeschrittenes Niveau und Bezug zu den oben aufgeführten Bereichen gegeben sein muss.

**Vereinfachte Anerkennung aus dem Auslandsstudium von bis zu 30 ECTS**

## 6.6 Zusammenfassung der formalen Randbedingungen der Modulwahl

- Pflichtbereich, 31 Credits: Beide Grundlagenmodule mit je 8 Credits sind Pflicht, ein Seminar (5 Credits) und ein Praktikum (10 Credits) sind ebenfalls Pflicht
- Grundlegender Wahlbereich (B1.1, B2.1, B3), mindestens 15 Credits: mindestens ein Modul aus jeder der drei Kategorien „Data Engineering“, „Data Analytics“ und „Data Analysis“ muss gewählt werden.
- Fortgeschrittener Wahlbereich (B1.2, B2.2), mindestens 25 Credits: entweder ein Forschungsprojekt oder ein Anwendungsprojekt aus der Kategorie „Data Engineering“, weitere 15 Credits Wahlmodule.

**Insgesamt** müssen

- aus dem Pflichtbereich (A), **31 Credits**



- zusammen aus dem Wahlbereich (B) und dem Wahlkatalog des Masterstudiengangs Informatik mindestens **53 Credits**
- aus dem Bereich Unterstützende Wahlmodule (D) **6 Credits**
- für die Master's Thesis (C) **30 Credits** erbracht werden.

## 6.7 Studien- und Stundenpläne

Im Folgenden ist exemplarisch ein Studienplan mit konkreter Auswahl der Module aus dem Pflichtbereich, dem Wahlbereich sowie den unterstützenden Wahlmodulen angegeben; der Studienbeginn ist Wintersemester.

### Erstes Fachsemester (Wintersemester)

<i>Modul</i>	<i>Abschnitt FPSO</i>	<i>Prüfung</i>	<i>CP</i>
Foundations in Data Engineering (IN2326)	A	Schriftl.	8
Query Optimization (IN2219)	B1.1	Schriftl.	6
Advanced Computer Networking (IN2097)	B1.2	Schriftl.	5
Data Analysis and Visualization in R (IN2339)	B2.1	Schriftl.	6
Applied Regression (MA4401)	B3	Schriftl.	5
<b>Gesamt:</b>			<b>30</b>

### Zweites Fachsemester (Sommersemester)

<i>Modul</i>	<i>Abschnitt FPSO</i>	<i>Prüfung</i>	<i>CP</i>
Foundations of Data Analysis (MA4800)	A	Schriftl.	8
Master-Praktikum (IN2106)	A	Projektarbeit	10
Englisch - Scientific Presentation and Writing C2	D1	Schriftl.	3
Algorithmen für Uncertainty Quantification (IN2345)	B2.2	Schriftl.	5
Anwendungsprojekt (IN2328) (Beginn der Arbeit) *	B1.2	Projektarbeit	4
<b>Gesamt:</b>			<b>30</b>

### Drittes Fachsemester (Wintersemester)

<i>Modul</i>	<i>Abschnitt FPSO</i>	<i>Prüfung</i>	<i>CP</i>
Master-Seminar (IN2107)	A	Wissenschaftliche Ausarbeitung	5
Machine Learning (IN2064)	B2.1	Schriftl.	8
Parallel Algorithms (IN2011)	B2.2	Schriftl.	8
Anwendungsprojekt (IN2328) (Hauptteil der Arbeit) *	B1.2	Projektarbeit	6
Social and Political Aspects of Data Science	D2	Schriftl.	3
<b>Gesamt:</b>			<b>30</b>

\* Das Anwendungsprojekt ist nicht an die Vorlesungszeiten gebunden, so dass dessen Workload von 10 CP auf das zweite und dritte Fachsemester gesplittet werden kann. Die Studienberatung schlägt regelmäßige Besprechungen zwischen dem Betreuer und Studierenden vor. Diese sind beispielhaft in den Stundenplan eingetragen, aber nicht verpflichtend, können also nach Bedarf zwischen dem Betreuer und Studierenden vereinbart werden.

### Viertes Fachsemester (Sommersemester)

<i>Modul</i>	<i>Abschnitt FPSO</i>	<i>CP</i>
Masterarbeit (IN2327)	§46	30
<b>Gesamt:</b>		<b>30</b>

### Credit-Bilanz

<b>Nr.</b>	<b>Bereich</b>	<b>Semester</b>	<b>Credits</b>
1	Pflichtbereich	1 - 3	31
2	Wahlbereich, unter Berücksichtigung der Randbedingungen: mindestens je ein Modul aus B1.1 B2.1, B3; IN2169 oder IN2338; mindestens 15 Credits aus B1.1, B2.1, B3; mindestens 25 Credits aus B1.2, B2.2	1 -3	53
3	Überfachliche Grundlagen	2	3

4	Social and Political Aspects of Data Science	3	3
5	Master's Thesis	4	30
	<b>Summe</b>		<b>120</b>

## 6.8 Rechtliche, ökonomische, ethische oder gender-orientierte Aspekte

Die Fakultät für Informatik, die Carl von Linde-Akademie und das MCTS bieten eine Reihe von Modulen mit Bezug zu Data Science an, die Studierenden des neuen Masterstudiengangs Einblicke in diverse rechtliche, ökonomische, ethische und gender-orientierte Aspekte geben können, z.B. Datenschutz, Datensicherheit, Datensparsamkeit. Module im Bereich überfachliche Grundlagen/Social and Political Aspects of Data Science können im Umfang von 6 Credits gewählt werden.

## 7 Organisatorische Anbindung und Zuständigkeiten

Der Studiengang "Data Engineering and Analytics" (DEA) wird unter Mitwirkung der TUM-Fakultät Mathematik angeboten. Die organisatorische und administrative Abwicklung erfolgt durch die Fakultät für Informatik, wobei angestrebt wird, gemeinsame Organe mit dem Partnerstudiengang "Mathematics in Data Science" zu schaffen.

Von außen wird das Programm sichtbar durch ein gemeinsames Webportal. Die Zulassung wird geregelt durch eine vereinheitlichte Bewerbungsprozedur, bei der die Bewerber angeleitet werden, sich für den passenden Studiengang zu entscheiden. Die Bewerbungen werden in enger Abstimmung evaluiert. Es wurde eine gemeinsame Studienkommission gebildet, die sich regelmäßig trifft, die Studiengänge bewertet und anpasst; die Studienkommission ist dabei offen für Beteiligung weiterer Fakultäten. Das Programm kann so als Keimzelle und Prototyp sowohl für eine interfakultäre Big Data-Initiative an der TUM als auch allgemein als Vorlage für integrative Studienprogramme fungieren.

Die Fachstudienberatung erfolgt über die Studiengangsverantwortlichen und den/die Studiengangskoordinator/in des Studiengangs. Diese wickeln darüber hinaus zusammen mit der Kommission für das Eignungsverfahren das Bewerbungsverfahren für den Masterstudiengang ab. Eine Hauptaufgabe ist hier die Beratung der Studieninteressierten, die sich für beide Studiengänge interessieren und hierfür qualifiziert sind. Die jeweiligen Qualifikationsprofile sind hier in enger Absprache zu erklären, so dass die Interessierten den Studiengang wählen, der ihren Interessen und Stärken am besten entspricht.

## 8 Ressourcen

### 8.1 Personelle Ressourcen

Die Fakultät für Informatik verfügt über eine neue Professur „Data Science“ und die bestehende Professur „Datenbanken“. Außerdem werden einige Module aus den Bereichen Bioinformatik, Datenbanken, High Performance Computing schon angeboten, so dass insgesamt kein neuer Bedarf an personellen Ressourcen entsteht.

Eine detaillierte Aufstellung der personellen Ressourcen für die einzelnen Module des Studienganges findet sich im Anhang 9.2.

## 8.2 Sachausstattung / Räume

Die Fakultäten für Informatik und für Mathematik verfügen über genügend Sachmittel (Tutorien, Lehr- und Lernmaterialien, Geräte, ...) zur ordnungsgemäßen Durchführung des Studiengangs DEA im Rahmen der angestrebten Studierendenzahlen.

Der Campus Stammgelände und der Campus Garching, inklusive Garching-Hochbrück der Technischen Universität München verfügen über genügend Räume für die ordnungsgemäße Durchführung des Studienganges.

Die Verwaltung der Studierenden erfolgt über das Campusmanagementsystem TUMonline. Es ermöglicht eine konsistente Führung der Studierendendaten, auf die sowohl das Immatrikulations- und Prüfungsamt als auch die Koordinatoren des Studiengangs zugreifen können. Auch ermöglicht es das fakultätsübergreifende Management der Leistungen der Studierenden.

Den Studierenden dient TUMonline als zentrales Portal zur Veranstaltungs- und Notenverwaltung. So ist es den Studierenden möglich, über TUMonline unter anderem auf das komplette für DEA relevante Lehrangebot zuzugreifen und somit Veranstaltungen für eines ihrer Module finden. Durch die Registrierung für eine Veranstaltung in TUMonline ist es einerseits den Dozenten problemlos möglich, mit einzelnen oder allen Teilnehmern der jeweiligen Lehrveranstaltung direkt in Kontakt zu treten, andererseits erhalten die Studierenden so einen aktuellen, individuellen Stundenplan. TUMonline dient weiter der persönlichen und direkten Bekanntgabe von Prüfungsergebnissen sowie dem unbürokratischen Abruf von Dokumenten wie Immatrikulationsbescheinigung und Leistungsnachweis.