

Big Data und Industrie 4.0

Natalie Gläser

23. August 2017

Hochschule für angewandte Wissenschaften Hamburg,
Dept. Informatik
Berliner Tor 7
20099 Hamburg, Deutschland
natalie.glaeser@haw-hamburg.de

Zusammenfassung

In dieser Ausarbeitung werden Big Data und Industrie 4.0 erläutert und deren Relevanz in der heutigen Zeit hervorgehoben. Die beide Begriffe werden im Verlauf der Arbeit in Zusammenhang gebracht. Dabei werden zwei Paper untersucht, die sich mit Big Data in der Industrie 4.0 befassen.

Schlüsselwörter: Big Data, Industrie 4.0, Smart Data, cyber-physische Systeme, Smart Factory, Internet der Dinge

1 Einführung

Für die Realisierung von Industrie 4.0 bedarf es einer Reihe von verschiedenen Technologien. Im Bereich eingebetteter Systeme gilt Deutschland bereits im Maschinen- und Automobilbau als führender Pionier. Durch cyber-physische Systeme (CPS) als Basistechnologie der vierten industriellen Revolution soll ein weiterer Meilenstein gelegt und Industrie 4.0 in die Praxis umgesetzt werden [KaLW11, S.2]. Im Zusammenhang mit Industrie 4.0 ist Big Data ein essentieller Themenbereich, denn in der Produktion entsteht eine Vielzahl von Daten, die gespeichert und sinnvoll verarbeitet werden muss. Big Data ist ohnehin durch das immer höhere Datenaufkommen in der heutigen Zeit ein populäres Thema, jedoch beschäftigen sich viele Forschungsorganisationen und Unternehmen primär mit von Menschen erzeugten Daten (beispielsweise für Verkaufsprognosen) und lassen maschinengenerierte Daten außer Acht [GKAE16, S.4]. Diese Ausarbeitung erörtert beide Themen und bringt diese in Relation.

2 Big Data

Der globale Datenbestand verdoppelt sich laut Prognosen alle zwei Jahre [MBKS17, S. 56]. Der Begriff Big Data kann als Synonym für große Datenmengen verstanden werden, welche beispielsweise im Internet durch soziale Netzwerke verursacht werden: mehrere Millionen Nutzer erzeugen eine Vielzahl an Daten, die gespeichert und verarbeitet werden müssen. Auch in der Wissenschaft sowie in Unternehmen verschiedenster Branchen ist Big Data ein gefragtes Thema [BuRM13, S. 63f.]. Unternehmen nutzen die Daten um beispielsweise das zukünftige Kaufverhalten der Kunden zu prognostizieren oder um vorbeugende Wartungsmaßnahmen vorzunehmen. Durch die Auswertung der Daten sollen kleinste Mängel erkannt werden, bevor es überhaupt zum Maschinenausfall kommt [MBKS17, S. 57]. In Anlehnung an Klein/Tran-Gia/Hartmann [KITrHa13, S. 319] wird Big Data auch als „ein abstrakter Oberbegriff für jegliche Art und Anzahl von Daten, die mit traditionellen Datenanalyseverfahren nicht mehr handhabbar sind und deshalb neuer Techniken und Technologien bedürfen“ beschrieben. Klassische RDBMS eignen sich nicht für sehr große Datenmengen, da die gegebene Leistungsfähigkeit nicht ausreicht und sie nicht auf unstrukturierte Daten ausgelegt sind. Für diesen Zweck sind NoSQL-Datenbanken besser geeignet [Müll14, S. 449ff.].

In der Literatur findet man häufig die Definition des 5-V- Modells um Big Data zu beschreiben. Die 5-V stehen für die englischen Bezeichnungen von

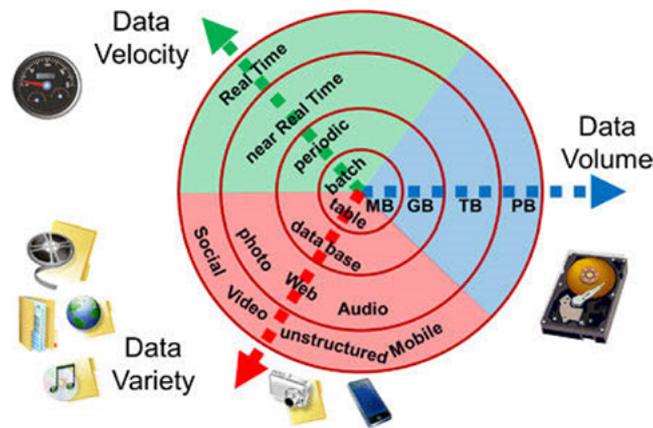


Abbildung 1: 3-V-Modell [KITrHa13, S.320]

Geschwindigkeit (velocity), *Zuverlässigkeit und Genauigkeit* (veracity), *Vielfalt* (variety), *Volumen* (volume) und *Wert* (value) [MBKS17, S. 57f.], wobei der Ursprung beim 3-V-Modell von Gartner (2011) liegt, dass nur für *Volumen*, *Geschwindigkeit* und *Vielfalt* steht [KITrHa13, S. 320].

Volumen steht für das stark ansteigende Datenaufkommen sowie für die Menge der Daten. Es ist jedoch nicht universell definiert, wie groß solche Datenmengen in Bezug auf Big Data sind. In der Literatur [MBKS17, S.57] werden Datenmengen ab einem Terabyte als groß eingestuft.

Die *Geschwindigkeit* bezieht sich sowohl auf die rasche Entstehung neuer Daten als auch auf die Verarbeitung und Speicherung der Daten. Hier gilt es als erstrebenswert, Daten in Echtzeit zu verarbeiten [VoHa16, S. 34], was zum Beispiel mit In-Memory- Datenbanksystemen umgesetzt werden kann.

Ein weiteres Hauptmerkmal von Big Data ist die *Vielfalt* der Daten, die Bezug auf die verschiedenen Datenquellen und auf die Struktur der Daten nimmt. Man unterscheidet hier zwischen strukturierten, halbstrukturierten, unstrukturierten oder polystrukturierten Daten [VoHa16, S.35]. Strukturierte Daten können in relationalen Datenbanken gespeichert werden. Halbstrukturierte Daten hingegen sind beispielsweise Emails, denn beim Inhalt der Email ist keine feste Struktur vorgegeben. Lediglich der Absender, Adressat und der Betreff können eine einheitliche Struktur aufweisen [KITrHa13, S.320]. Unstrukturierte Daten sind zum Beispiel Audiodateien oder Videodateien, die nicht in traditionellen Datenbanken gespeichert werden, sondern durch spezielle Big-Data Analysen verarbeitet werden. Als polystrukturierte Daten werden zusammengeführte Daten mit unterschiedlichen Strukturen aus verschiedenen Datenquellen bezeichnet.

Der Faktor *Zuverlässigkeit* der Daten als ergänzende Eigenschaft zum 3-V-

Modell referenziert auf die Qualität der Daten. Es könnte die Gefahr bestehen, dass Daten aus unzuverlässigen Quellen stammen oder durch ungeeignete Analyseverfahren die Auswertung der Daten zu einem unbrauchbaren Ergebnis führen.

Ebenso wird der *Wert* der Daten als eine zusätzliche Eigenschaft von Big Data betrachtet. Hierbei handelt es sich um den unternehmerischen Mehrwert, der aus den Daten gewonnen werden kann. Durch die Analyse von halbstrukturierten und unstrukturierten Daten können für ein Unternehmen wichtige Erkenntnisse gewonnen werden, die eventuell in unternehmerische Entscheidungen mit einfließen.

3 Industrie 4.0

Das Bundesministerium für Bildung und Forschung, welches den Begriff als solches geprägt hat, beschreibt Industrie 4.0 unter anderem mit folgenden Worten[PlatoJ]:

„Technische Grundlage [für Industrie 4.0] [...] sind intelligente, digital vernetzte Systeme, mit deren Hilfe eine weitestgehend selbstorganisierte Produktion möglich wird: Menschen, Maschinen, Anlagen, Logistik und Produkte kommunizieren und kooperieren in der Industrie 4.0 direkt miteinander. Produktions- und Logistikprozesse zwischen Unternehmen im selben Produktionsprozess werden intelligent miteinander verzahnt, um die Produktion noch effizienter und flexibler zu gestalten.“

Um den Begriff Industrie 4.0 weitergehend zu erläutern, werden zunächst die vorangegangenen industriellen Revolutionen aufgeführt[Ligg17, S.210f.]. Die erste industrielle Revolution Ende des 18. Jahrhunderts ist durch den Einsatz von Maschinen gekennzeichnet. Die zweite industrielle Revolution am Anfang des 20. Jahrhundert brachte unter anderem die Massenproduktion mit sich. Durch die Arbeitsteilung haben Produktionsverfahren ein völlig neues Level erreicht. In der dritten industriellen Revolution in den siebziger Jahren konnten durch den Einsatz von Computern Arbeitsschritte automatisiert und Prozesse optimiert werden [AnHä17, S.3]. In der uns noch bevorstehenden vierten industriellen Revolution geht es unter anderem um die massenindividualisierte Produktion. Das Produkt soll kostengünstig und dennoch individuell auf den Kunden abgestimmt produziert werden. Darüber hinaus ist die autarke Instandhaltung von Produktionsmaschinen ein Thema von Industrie 4.0. Autonome Maschinen sollen eine effizientere und kostengünstigere Produktion ermöglichen, da beispielsweise Ausschüsse durch frühzeitige Fehlerbehebung reduziert werden. Darüber hinaus soll eine Steigerung der Arbeits-

qualität möglich sein, indem Mitarbeiter nicht mehr als einfache „Maschinenbediener“ sondern als „Koordinator“ agieren und durch die Autonomie der Maschinen die Work-Life-Balance fördern [Kage17,S.240]. Das Fundament dieser Entwicklung bilden das *Internet der Dinge*. Das Internet der Dinge beschreibt die „Vernetzung von Objekten aller Art [. . .], deren Hauptzweck nicht auf abstrakte Transformationen sondern auf physikalisch-mechanische Aufgaben ausgerichtet ist“ [Bous17, S.5], denn für die Realisierung von Industrie 4.0 braucht es Maschinen, Lagersysteme und Betriebsmittel die miteinander vernetzt und in der Lage sind, miteinander zu kommunizieren. Diese Maschinen werden cyber- physischer Systeme (CPS) genannt und gelten als Basis von Industrie 4.0. Solche Systeme nehmen mittels Sensoren sämtliche Daten ihrer Umwelt auf, speichern diese und können aufgrund dieser Datengrundlage entsprechende physikalische Eingriffe vornehmen [Voge14, S.37f.].

Die Vorstellung von Industrie 4.0 in der Praxis wird vom Bundesministerium für Bildung und Forschung am folgenden fiktiven Beispiel erläutert [BMBFoJ]:

In einer Seifenspender- Fabrik sind alle Maschinen, Lagersysteme sowie Betriebsmittel in der Lage, sich gegenseitig zu steuern. Auf eine zentrale Steuerung wird verzichtet. Die Rohlinge erhalten einen Smart- Tag, welches alle Informationen für den Produktionsvorgang und der Auslieferung enthält. Anhand des Smart- Tags erkennen die Maschinen, wie das Produkt ausgestattet werden soll. Das kann die Farbe der Seife, die Art der Seifenspender Pumpe oder die Sprache des Etiketts sein. Darüber hinaus werden Nachbestellungen von Rohstoffen automatisiert in Echtzeit veranlasst. Der Kunde hat die Möglichkeit, den gesamten Vorgang seiner Bestellung über das Internet zu verfolgen. So wird eine höhere Transparenz gegenüber dem Kunden gewährleistet.

4 Big Data und Industrie 4.0

Durch die Nutzung von CPS entsteht ein großes Datenaufkommen, weshalb Big Data in Bezug auf Industrie 4.0 ein relevantes Thema darstellt. Diese Daten müssen nach ihrer Erzeugung gespeichert und verarbeitet werden. Eine Sammlung der Daten ist zwingend notwendig, da CPS eine Datenbasis brauchen, auf die sie bei ihren Handlungsentscheidungen zurückgreifen können. Dies setzt voraus, dass die gesammelten Daten analysiert werden. Die aus der Analyse gewonnenen Informationen werden auch *Smart Data* genannt [Kage17, S.237f.].

Lee/Kao/Yang [LeHuYa14, S.5ff.] zeigt ein grobes Konzept für ein cyber-

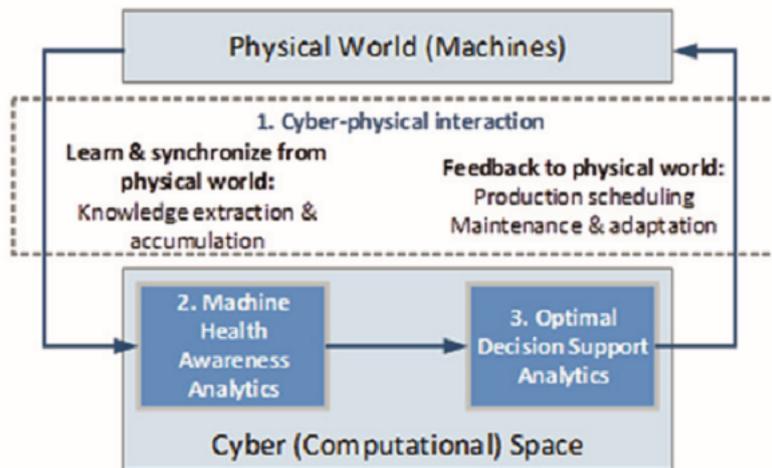


Abbildung 2: Cyber-physical system framework for self-aware and self-maintenance machines [LeHuYa, S.5]

physical System Framework für autonome Maschinen in Bezug auf die selbstständige Wartung (Abbildung 2). Die *Physical World* beinhaltet mehrere Maschinen, die über Reglerparameter, Zustandsüberwachung und eine digitalisierte Maschinenperformance verfügen. Das Format der Daten der physikalischen Welt wird definiert, damit eine entsprechende Datenerfassung erfolgen kann. Die Daten werden gesammelt und das Wissen über den Maschinenverschleiß für die Selbsteinschätzung zusammengefasst (*Machine Health Awareness Analytics*). Dieser Vorgang wird anhand von Abbildung 3 näher erläutert. Man geht davon aus, dass mehrere Maschinen mit der gleichen Funktion auf die selbe Datenbasis zurückgreifen. Die gesammelten Daten über das Maschinenverhalten werden chronologisch und von Maschine zu Maschine verglichen. Anschließend prüfen Algorithmen, ob die Informationen einem Cluster zugeordnet werden können. Ist dem so, wird das entsprechende Cluster aktualisiert. Können mehrere Informationen keinem Cluster zugeordnet werden, werden die Informationen über das Verhalten der Maschinen so lange gesammelt, bis genügend Informationen vorhanden sind, um ein neues Cluster zu erstellen. Mit Hilfe des erlangten Wissens wird ein entsprechender Handlungsbedarf an die physikalische Schicht weitergegeben (*Optimal Decision Support Analytics*).

In [GKAE16] wird ebenfalls ein Konzept einer Architektur für Big Data in der Industrie 4.0 beschrieben, welches im Vergleich zum bereits genannten Framework eine höheren Abstraktionsebene aufweist. Es wird die These aufgestellt, dass die Komplexität von Big Data in der Industrie 4.0 Unter-

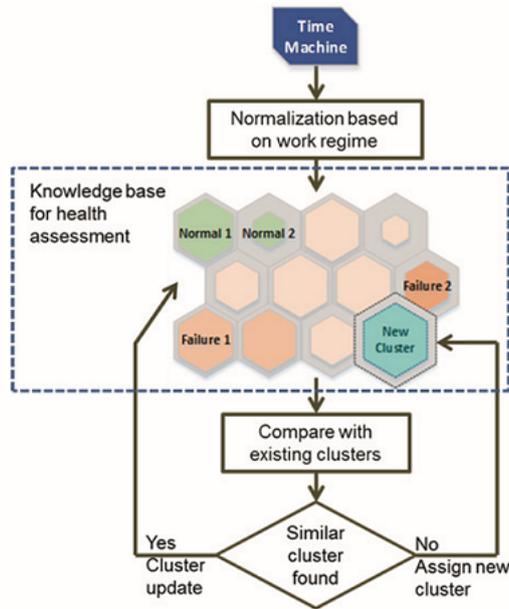


Abbildung 3: Adaptive learning for machine clustering [LeHuYa, S.6]

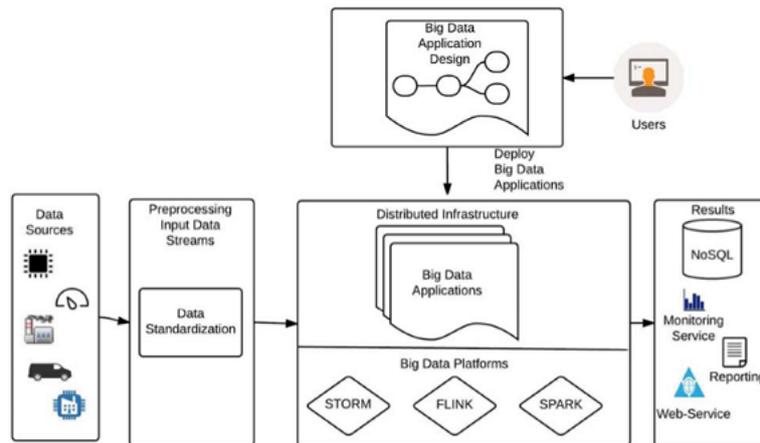


Abbildung 4: Architectural Conceptual Framework [GKAE16, S.432]

nehmen bei fehlendem Expertenwissen hindern könnte, Big Data Technologien einzusetzen. Das genannte Konzept soll Softwarearchitekten ermöglichen, auch bei geringem Erfahrungsschatz eigene Big Data Use Cases für Industrie 4.0 Anwendungen zu entwickeln. Wie in Abbildung 4 zu erkennen, besteht das Framework aus vier Modulen. Im *Big Data Application Design Modul* werden die Applikationen als gerichteter Graph dargestellt. Die Datenströme zwischen den Anwendungen werden als Kanten abgebildet. Die Applikationslogik entsteht durch das Verbinden der Knoten, welche die Programmstrukture darstellen. Im *Pre Processing Input Data Streams- Modul* werden Daten mit unterschiedlicher Struktur aus *Data Sources* für die Weiterverarbeitung standardisiert. Das *Distributed Infrastructure- Modul* enthält unterschiedliche Big Data Plattformen wie Storm, Spark oder Flink, damit die Applikationen für jeden Anwendungsfall die passende Plattform nutzen können, da je nach Fall die Anforderungen an die Plattform variieren können. Ebenso entscheidet die Applikationslogik über die Auswahl der Plattform. Im *Distribution of Results- Modul* werden die Ergebnisse auf sämtlichen Kanälen in unterschiedlichen Formen an die zuständigen Mitarbeiter weitergegeben. Diese Ergebnisse können auch als Input verwendet werden und zur Optimierung der Prozesse beitragen.

5 Fazit und Ausblick

Big Data ist für die Entwicklung von Industrie 4.0 ein wichtiges Thema, da das Aufkommen von große Datenmengen mit der Verwendung von CPS unvermeidbar ist. Die korrekte Verarbeitung der Daten ist eine Grundvoraussetzung für ein funktionierendes Gesamtsystem in der Industrie 4.0. Die Paper zeigen, wie Big Data und Industrie 4.0 zusammenhängen. Außerdem werden grobe Ansätze für die Realisierung vorgestellt, jedoch werden geeignete Big- Data- Technologien für maschinengenerierte Daten nicht weiter thematisiert. Vielmehr handelt es sich hierbei um Frameworks, die sich unter Umständen pauschalisieren und auch auf andere Bereiche abbilden lassen. Um das Thema Big Data in der Industrie 4.0 weiter zu vertiefen, könnte unter anderem die Frage nach geeigneten Big Data- Technologien für die Erfassung und Verarbeitung maschinengenerierter Daten ein weiterer Aspekt sein.

Literatur

- [AnHä17] Andelfinger, V. P., Hänisch, T.: Industrie 4.0: Wie cyber-physische Systeme die Arbeitswelt verändern. Springer, Wiesbaden 2017, S. 2-4
- [BMBFoJ] Bundesministerium für Bildung und Forschung: Industrie 4.0. <https://www.bmbf.de/de/zukunftsprojekt-industrie-4-0-848.html> , Abruf am 17. Mai 2017
- [Bous17] Bousonville, T.: Logistik 4.0: Die digitale Transformation der Wertschöpfungskette. Springer, Wiesbaden, 2017
- [BrSt16] Brecht, L., Stelzer B.: Big Data im Kontext von Industrie 4.0: Eine Technologievorausschau anhand IT-gestützter bibliometrischer Analyse und Szenariotechnik. ITOP- Schriftenreihe Nr.7, 2016, <https://oparu.uni-ulm.de/xmlui/handle/123456789/4247>, Abruf am 10. Mai 2017
- [BuRM13] Buhl, H., Röglinger, M., Moser F.: Big Data: Ein(ir-)relevanter Modebegriff für Wissenschaft und Praxis? Wirtschaftsinformatik, Volume 55, Issue 2, 2013
- [GKAE16] Gökalp, M.O., Kayabay, K., Akyol M.A., Eren P.E., Kocyigit, A.: Big Data for Industry 4.0: A Conceptual Framework. 2016 International Conference on Computational Science and Computational Intelligence. 2016, S. 432-433
- [KlTrHa13] Klein, D., Tran-Gia, P., Hartmann, M.: Big Data. Informatik Spektrum, 2013, Band 36, S. 319-323
- [Kage14] Kagermann, H.: Industrie 4.0 und Smart Services. In: Brenner, W.; Hess. T.: (Hrsg.): Wirtschaftsinformatik in Wissen, Springer, Berlin, 2014, S. 244-247
- [Kage17] Kagermann, H.: Chancen von Industrie 4.0 nutzen. In: Bauernhansl, T.; ten Hompel, M.; Vogel-Heuser, B.(Hrsg.): Handbuch Industrie 4.0 Bd.4- Allgemeine Grundlagen, Berlin, 2017
- [KaLW11] Kagermann, H.; Lukas, W.-D.; Wahlster, W.: Industrie 4.0: Mit dem Internet der Dinge auf dem Weg zur 4. industriellen Revolution. VDI- Nachrichten, 2011, Nr.13
- [LeHuYa14] Lee J., Hung-An K., Yang, S.: Service Innovation and smart analytics for Industry 4.0 and big data environment. Product Services System and Value Creation. Proceedings of the 6th CIRP Conference on Industrial Product- Service Systems. 2014, S. 4-6

- [Ligg17] Liggesmeyer, P.: Alles 4.0! Oder manchmal doch 3.5? Informatik Spektrum, 2017, Band 40, Heft 2
- [MBKS17] Mertens, P.; Bodendorf, F.; König, W.; Schumann, M.; Hess, T.; Buxmann, P.: Grundzüge der Wirtschaftsinformatik. 12. Auflage, Springer, Berlin, 2017, S. 56-59
- [Müll14] Müller, S.: Die neue Realität: Erweiterung des Data Warehouse um Hadoop, NoSQL & Co. HMD Praxis der Wirtschaftsinformatik: Big Data, Volume 51, Issue 4, 2014
- [PlatoJ] Plattform Industrie 4.0: Was ist Industrie 4.0? <http://www.plattform40.de/I40/Navigation/DE/Industrie40/WasIndustrie40/was-ist-industrie-40.html>, Abruf am 17. Mai 2017
- [Voge14] Vogel-Heuser, B.: Herausforderungen und Anforderungen aus Sicht der IT und der Automatisierungstechnik. In: Bauernhansl, T.; ten Hompel, M.; Vogel-Heuser, B. (Hrsg.): Industrie 4.0 in Produktion, Automatisierung und Logistik: Anwendung-Technologien-Migration, 2014, S. 37-48
- [VoHa16] Volk, M.; Hart, V.-W.: Big-Data: Bestimmung der Big-Data-Artigkeiten von Projekten. In: Schenk, M.: (Hrsg.): Digitale Methoden und Modelle im Anlagenbau, Fraunhofer, 2016, S. 32-40