

Big Data in der Produktion – Teil 1

Christian Zwickl-Bernhard

www.mcp-zwickl-bernhard.at

November 2016

Inhaltsangabe

Was ist Big Data?	1
Steigende Datenvolumina	2
Unterschied zwischen Big Data und klassischen Datenbank-Anwendungen	3
Zentrale Begriffe	4
Anforderungen an die Unternehmen	5
Die Möglichkeiten von Big Data Anwendungen	6
Neue Berufsbilder	6

Was ist Big Data?

Industrie 4.0, das Schlagwort moderner Produktionstechnik kann im Kern als die Verschmelzung digitaler und physischer Prozesse aufgefasst werden. Einzelne Produktionsschritte werden systematisch durch computer- bzw. datengestützte Prozesse ersetzt oder zumindest begleitet. Produkte, Maschinen, Anlagen, Werkzeuge, ja selbst Gebäude stellen Daten in unterschiedlicher Form bereit und werden gleichzeitig in die Lage versetzt auf diese Daten zu reagieren bzw. diese auch zu verarbeiten. Das Ziel von Industrie 4.0 kann in dieser Hinsicht als die weitgehende Digitalisierung der Wertschöpfungskette sowie deren Optimierung im Hinblick auf Individualität und Flexibilität der Produktion aufgefasst werden. Der Zusammenhang zwischen den Konzepten von Industrie 4.0 und der sich daraus ergebenden Menge an Daten wird damit verständlich und erklärbar. Moderne Produktionsmethodik ist ohne die Bewältigung der sich aus den Konzepten von Industrie 4.0 ergebenden Datenmengen undenkbar, ja Industrie 4.0 ist in seiner Umsetzung eigentlich ein weitgehend von Software und damit von Daten bestimmtes Konzept.

Der Unterschied zu bisherigen Konzepten liegt bedingt durch die immer weitergehende Durchdringung der gesamten Prozesskette mit Daten in der Zunahme an Volumen, der

stärkeren Unterschiedlichkeit der zu bearbeitenden Datenstrukturen sowie in der steigenden Frequenz des Datenaufkommens. Man spricht in dem Zusammenhang von den „3V“, also Volume, Variety und Velocity. Sensoren generieren riesige Datenmengen (Volume), die in sehr kurzen Zeitabständen auf unterschiedliche Charakteristiken analysiert werden müssen (Velocity) und in Beziehung zu vollkommen unterschiedlich strukturierten Daten (Variety) zu setzen sind. Gerade im Produktionsbereich besteht hier in vielen Fällen auch die Forderung nach einer Echtzeitverarbeitung der Daten.

Die Unterschiede zu klassischen Datenbank-Anwendungen¹ sind nur graduell, sie erfordern dennoch vollkommen neue Konzepte, Ansätze und Werkzeuge. Und auch die BIG-DATA spezifischen Berufsbilder beginnen sich langsam abzuzeichnen.

Big Data kann damit als die Verarbeitung sehr großer und unterschiedlich strukturierter Datenmengen in Echtzeit verstanden werden. Mit anderen Worten ist Big Data die *wirtschaftlich sinnvolle Gewinnung und Nutzung entscheidungsrelevanter Erkenntnisse aus qualitativ vielfältigen und unterschiedlich strukturierten Informationen, die einem schnellen Wandel unterliegen und in bisher unbekanntem Umfang anfallen* (BITKOM, 2012). Dies erfordert neue Technologien, neue Werkzeuge und adaptierte Modellansätze. Eine Herausforderung für jedes Unternehmen, eine Herausforderung die auch Auswirkungen auf die unternehmensinterne Organisation hat. Daten stellen neben der Arbeitskraft, dem Kapital sowie den Rohstoffen den vierten und in Zukunft immer wichtiger werdenden Produktionsfaktor dar. Big Data ist das Management dieses neuen Produktionsfaktors. Eine Herausforderung der sich jedes zukunftsorientierte Unternehmen früher oder später stellen muss.

Steigende Datenvolumina

Im Zuge der Implementierung von Industrie 4.0 Konzepten gibt es zahlreiche Gründe für die zunehmende Menge und Komplexität der anfallenden Daten. Beinahe alle Technologien, die im Rahmen von Industrie 4.0 Anwendung finden, führen zu steigenden Datenvolumina. Zunächst nimmt die Intelligenz der verarbeitenden Produkte massiv zu. Dies ist direkt mit einem Anstieg der Datenmenge verbunden, genauso wie die Anforderung Individualität (Stichwort „Losgröße 1“) der Produktion mit gleichzeitiger Optimierung der Produktivität zu verbinden. Das Konzept der Digitalisierung des gesamten Produktionsverlaufes trägt zur Zunahme ebenso bei wie die generelle Forderung nach schnelleren Reaktionszeiten, auf Kundenbedürfnisse ebenso wie auf allfällige Störungen im Produktionsablauf (Stichwort „Predictive Maintenance“).

¹ Vgl. Abbildung 1

Das Konzept *Ubiquitäres Computing* führt über die zunehmende Anzahl von Sensoren zu weiteren Optimierungs- und Gestaltungsmöglichkeiten und zu Daten, die zumeist in Echtzeit verarbeitet werden müssen. Hinzu kommt dass sowohl Produkt- als auch Produktionsdaten als Optimierungsquellen für den Betriebszustand, die aktuelle Produktionsleistung oder die Produktionsqualität genutzt werden können. Die Kommunikation einzelner Maschinen untereinander sowie der zunehmende Einsatz von Sensoren zur Überwachung der Abläufe führen ihrerseits zu höheren bzw. dichteren internen Datenflüssen.

Aber auch der generelle Trend zu einem höheren Anteil an Dienstleistung führt sowohl innerhalb des Unternehmens als auch an der Schnittstelle zu Kunden zu einem höheren Datenaufkommen, das gemanagt, protokolliert und verarbeitet werden muss.

Um die genannten Anforderungen abzudecken sind auch entsprechende Speicherkonzepte notwendig, wie zum Beispiel die Cloud Technologie, ebenfalls ein wesentlicher Eckpfeiler von Industrie 4.0.

Unterschied zwischen Big Data und klassischen Datenbank-Anwendungen

Obwohl der Unterschied graduell ist, gibt es doch einige grundlegend neue Themen im Umfeld von Big Data Anwendungen.

Zunächst sind es die „drei V“ die den wesentlichen Unterschied ausmachen. Die Datenmengen die verarbeitet werden können sind wesentlich höher („Volume“). Man spricht in dem Zusammenhang auch von der Tera-, Peta- oder Zettabyte-Technologie.² Velocity bedeutet in der Praxis Reaktionsgeschwindigkeit in Echtzeit, bevor ein Mensch noch wahrnehmen kann, was da passiert bzw. passiert ist. Variety bedeutet einerseits die mögliche Vielfalt an unterschiedlichen Datenquellen sowie auch die technische Vielfalt der Daten selbst. Texte, Audioaufzeichnungen oder Videos müssen gesammelt, strukturiert ausgewertet und in unterschiedlichen Beziehungen zueinander gesetzt werden.

Klassische Datenbankanwendungen basieren zumeist auf homogenen Datenstrukturen, klassischen Transaktionskonzepten sowie auf SQL-basierenden Abfragewerkzeugen. Big Data Anwendungen können auch unstrukturierte Daten mit Hilfe des Map-Reduce Ansatzes verarbeiten und dies basierend auf Rechner-Cluster in extrem hohen Verarbeitungsfrequenzen.

Aus dem Gesagten ergibt sich auch eine wesentliche Konsequenz für die organisatorische

² 1 Zettabyte = 10¹² Gigabyte

Aufstellung in Unternehmen. Die IT-Abteilungen können nicht mehr unabhängig von Fachabteilungen Lösungen bereitstellen, sondern sie werden eingebunden in ein Team aus Daten-Architekten zum Bereitsteller der notwendigen Architektur und IT-Landschaft. Die Verantwortung für die Big Data Anwendungen werden zunehmend dem CIO oder einem CDO („Chief Data Officer“) übertragen.

Zentrale Begriffe

Big Data wird bisher zumeist für Marketing, die Pflege von Kundenbeziehungen, die Analysen des Verhaltens der Kunden, die Beobachtung von Veränderungen am Markt, etc. eingesetzt. Zunehmend findet Big Data aber auch den Eingang in die konkrete Produktionslandschaft. Hier gelten jedoch besondere Anforderungen bzw. Rahmenbedingungen.

Die neuen Technologien beziehen sich im Besonderen auf folgende Themen:

- Verteilte Verarbeitung der Daten
- Einsatz von In-Memory Technologien (Hauptspeicher-residente Datenbanken)
- Bereitstellung von Tools, die für die Analyse großer Datenmengen geeignet sind.

Es werden in den Anwendungen Cockpits für Auswertungen auf Management-Ebene ebenso bereitgestellt wie modellgestützte, auch in Echtzeit verfügbare Analysemethoden. Eine besondere Herausforderung stellt dabei die Einbeziehung unstrukturierter Daten dar (Stichwort „Variety“).

Die zentralen Begriffe in diesem Zusammenhang sind CEP, Data Discovery bzw. Data Mining, HADOOP, Map reduce Ansatz, NoSQL, Visual Analytics, DAM sowie In-Memory Technologien.³ Hinzu kommen Themen wie maschinelles Lernen, NLP sowie die notwendige Beherrschung von Skript-Sprachen.

HADOOP stellt die zentrale Technologie für Big Data Anwendungen dar. HADOOP ist ein frei verfügbares, in JAVA geschriebenes Framework für verteilt arbeitende Software. Im Wesentlichen besteht HADOOP aus einem verteilten Dateisystem als Basis (HDFS – „Hadoop Distributed File System“) sowie einem auf diesem Filesystem aufbauenden Algorithmus zur Parallelisierung von Analyse- und Auswerteprozessen. Dieser basiert auf dem MapReduce Ansatz. Dabei werden einfach ausgedrückt Datenpakete an unterschiedliche Rechner im Netzwerk („Hadoop-Cluster“) verteilt und dort parallel bearbeitet. Nach erfolgter Abarbeitung werden die Ergebnisse wieder an den Server gesendet und dort weiter verarbeitet. Diese beiden Schritt „map“ und „reduce“ können

³ Vgl. Tabelle 2

vom Anwender gesteuert werden, dazwischen liegt noch der Shuffle-Schritt zur Optimierung der Ergebnisse. Der Shuffle-Schritt kann zusätzlich noch auf Fehler im Hadoop-Cluster, etwa den Ausfall von Knoten, reagieren und diesen Fehler ausgleichen.

Data Mining oder CEP („Complex Event Processing“) sind klassische Technologien, die auch im Big Data Umfeld zum Einsatz kommen. CEP wird zur Analyse der Ergebnisse eingesetzt, um außergewöhnliche Werte (Ausreißer) aus dem Datenstrom zu erkennen, Trends zu identifizieren sowie Summen, Minima, Maxima oder dgl. zu bilden und an die Steuerungs- und Auswerteebene zu kommunizieren.

Anforderungen an die Unternehmen

Entscheidet sich ein Unternehmen für den Einsatz von Big Data, so kommen grundlegend neue Herausforderungen auf das Unternehmen zu. Neben der Bereitstellung und Implementierung der notwendigen IT-Infrastruktur (Hadoop- bzw. Rechner-Cluster) bildet die notwendige klar formulierte Zielsetzung für das Big Data Projekt den Ausgangspunkt, inklusive der Festlegung einer mit erster Priorität umzusetzenden Pilotanwendung.

Auf Managementebene wird das Team aus den beteiligten Fachabteilungen zusammengesetzt. Hier entstehen mittelfristig vollkommen neue Berufsbilder. Unter fachlicher Begleitung eines Big Data Spezialisten kann dann im Weiteren das Vorgehensmodell festgelegt werden. Dies beinhaltet auch ein Konzept welche Daten, welche vorhandenen Datenbankanwendungen bzw. welche klassischen Data-Mining Systeme in die neue Anwendung eingebunden werden sollen.

Erst auf Basis dieses Vorgehensmodells kann die Auswahl der Tools erfolgen. Hierbei sind zahlreiche Aspekte zu berücksichtigen: Gewünschte Offenheit der Anwendungen, Open Source JA/NEIN, vorhandene Schnittstellen-Module, Performance Anforderungen, Lauffähigkeit auf dem bereitgestellten Rechner-Cluster, etc.

Der nächste Schritt besteht in der Schulung des Teams (die Basis bildet ein Ausbildungskonzept), damit im gesamten Team ein weitgehend gleiches Know-how zu allen notwendigen und eingesetzten Werkzeugen und Verfahren besteht.

Vor Implementierung der Pilotanwendung sind auch die Verantwortlichkeiten für Datenschutz, Datensicherheit und das Datenmanagement zu regeln und klar festzulegen. Spätestens an dem Punkt zeigt sich auch, dass die klassische IT und die Produktions-IT in ein Big Data Projekt nur gemeinsam erfolgreich umsetzen können. Die bisher in vielen Fällen noch bestehenden Schranken zwischen den Abteilungen müssen spätestens hier aufgebrochen werden. Ein gemeinsames Verständnis, eine gemeinsame Sicht und ein übergreifendes Know-how bilden die Grundvoraussetzungen für ein erfolgreiches Big Data

Projekt.

Die Möglichkeiten von Big Data Anwendungen

Mit dem Einsatz grundlegend neuer Technologien und Konzepte im Rahmen von Big Data Projekten sind sich auch vollkommen neue Anwendungsmöglichkeiten verbunden. Die zeitnahe Einbeziehung von Produkt- und Produktionsdaten erlaubt die dynamische Anpassung von Geschäfts- und Produktionsprozessen. Im herkömmlichen Business Process Management (BPM) sind die Geschäftsprozesse statisch entworfen. Damit ist gemeint, dass die Geschäftsprozesse zumeist den Vorgaben zum Zeitpunkt der Entwicklung der Anwendungen entsprechen. Die zukünftigen BPM-Anwendungen werden die Möglichkeit bieten, die aus den aktuellen Produktionsprozessen gesammelten Daten zeitnah aufzubereiten und in das BPM einzubeziehen. Man spricht dann von iBPMs („Intelligent Business Process Management Suites“).

Die Einbeziehung neuer Datenquellen, wie Social Media Daten, Daten von mobilen Endgeräten oder CEP-Anwendungen erweitern die Grundlage von Marketing- und Vertriebsentscheidungen. Im Rahmen von Business Activity Monitoring (BAM) werden Daten visuell aufbereitet, etwa im Rahmen von Business-Dashboards.

Generell dienen Big Data Anwendungen auch zur Beherrschung der zunehmenden Komplexität bei den Produktions- und Geschäftsprozessen.

Durch die höhere Dichte und verbesserte Qualität der vorliegenden Daten ergeben sich neue Geschäftsmodelle, Entscheidungen fallen auf allen Ebenen (Management, Produktion, ..) auf Basis qualifizierter Datenstrukturen. Man spricht in diesem Fall von Big Data basiertem *Decision Support*. Die Transparenz erhöht sich und bei geschickter Struktur der Big Data Anwendung entstehen auch Dokumentation sowie Entscheidungspapiere als Nebenprodukte.

Gerade auch im Bereich produzierender Unternehmen können Lagerkosten reduziert werden, die Produktplanung kann optimiert und verbessert werden. Durch Big Data basiertes Predictive Maintenance lassen sich Produktionsausfälle minimieren bzw. ungeplante Maschinen- oder Anlagenstillstände vermeiden.

Neue Berufsbilder

Sowohl innerhalb eines Unternehmens als auch auf dem externen Consulting-Markt bilden sich im Rahmen von Big Data Projekten oder Anwendungen neue Berufsbilder. Man spricht beispielsweise vom *Chief Data Officer* (CDO) oder vom *Data Scientist*. Der Daten-

Architekt ist für die Gestaltung des Datenmodells zuständig und dies erfordert Kenntnisse aus dem Big Data Umfeld ebenso wie Know-how zu den einzelnen Produktionsschritten. Der *Data Artist* ist für die zumeist grafische Aufbereitung bzw. Visualisierung der Ergebnisse zuständig. Big Data Programmierer sollten neben JAVA auch die verwendeten script-Sprachen sowie die zahlreichen eingesetzten Tools (HADOOP, HDFS, ..) beherrschen. Aber auch Know-how zu den eingesetzten Schnittstellen-Konzepten ist unerlässlich.

Erfolgreiche Big Data Teams sind bereit neue Wege zu gehen und neue Ansätze zu verfolgen, sie nehmen ungewohnte Konzepte des maschinellen Lernens auf und setzen diese um, sie haben durchgängig hohe analytische Fähigkeiten, etwa bei der Einbindung unstrukturierter Daten. Je breiter das Know-how in einem Team ist, desto eher ist sichergestellt, dass die entstehenden Big Data Anwendungen auch den Herausforderungen genügen. „Vertikale“ Datenwissenschaftler wie Informatiker, Datenbank-Spezialisten oder Softwareingenieure sollten in den Teams mit Wirtschaftsanalysten oder Statistikern, den sogenannten „horizontalen“ Datenwissenschaftlern kombiniert zum Einsatz kommen. Hinzu kommen noch Experten aus dem Produkt- bzw. Produktionsbereich.

Datenwissenschaftler sind eine Art Mischung aus Hacker, Analyst, Kommunikator und vertrauenswürdigem Berater. Zusätzlich notwendige Eigenschaften sind die Fähigkeit zu assoziativem Denken sowie hohe soziale Kompetenz.

Dass Unternehmen ihrerseits auch einen Beitrag leisten müssen, diese Fachkräfte einerseits zu finden, weiter auszubilden und dann auch an das Unternehmen zu binden, liegt in Anbetracht der stark steigenden Nachfrage nach Big Data Fachkräften auf der Hand.