

Produktionsdatenanalyse in der Industrie



von Dipl.-Ing (FH) Markus Woehl (VIDEC Data Engineering GmbH, www.videc.de)

Produktionsdatenanalyse in der Industrie

1. Einleitung

"Daten sind das neue Gold" ist eine viel bemühte Phrase in den Medien. Im Folgenden werden wir Herangehensweisen in verschiedenen Bereichen der industriellen Datenverarbeitung betrachten. Im Speziellen werden wir dabei den Fokus auf Daten im Produktionsumfeld legen.

Immer wenn eine voll- oder halbautomatische Produktion abläuft, entstehen neben den Produkten auch Daten zum Produktionsumfeld. Zumeist sind diese Daten technisch mess- und numerisch darstellbar. Sie werden in der Regel von den Steuerungssystemen der Herstellungsaggregate mit erzeugt bzw. bereitgestellt. Das können Herstellungsparameter wie Temperatur, Druck, Viskosität und vieles andere mehr sein.

2. Infrastruktur

Im ersten Schritt werden die Daten dazu verwendet, den aktuellen Zustand der Produktion dem Anwender zur Verfügung zu stellen. Das geschieht zumeist über HMI SCADA (**H**uman **M**achine Interface **S**canning, **A**larming and **D**ata **A**cquisition) Systeme, die konzipiert sind, den aktuellen Zustand einer Maschine oder Produktionsanlage dem Anwender möglichst übersichtlich darzustellen.

Aber nicht nur die Erfassung und Darstellung verdient Aufmerksamkeit. Auch die Sicherheit der Produktionsdaten und der Schutz vor Cyberangriffen müssen gleich von Anfang an berücksichtigt werden. Zu diesem Themenkomplex steht ein separates Whitepaper ("Sichere Produktion") auf unserer Webseite zur Verfügung.

Inhalt

1. <u>Einleitung</u>	1
2. <u>Infrastruktur</u>	1
3. <u>Datenerfassung</u>	2
4. <u>Datenspeicher und Aggregation</u>	2
1. <u>Texte</u>	4
5. <u>Berechnungen und Kennzahlen</u>	4
6. <u>Auswertung</u>	4
1. <u>Darstellungen</u>	4
2. <u>Trends und Charts</u>	5
3. <u>Dashboards</u>	5
4. <u>Berichte</u>	5
5. <u>Alarmer und Ereignisse</u>	6
7. <u>Verfügbarkeit und Datensicherung</u>	6
8. <u>Regulierte Anwendungsfälle</u>	6
9. <u>Fazit</u>	7

3. Datenerfassung

Der Datenursprung ist der Sensor oder das Messgerät, welches den Parameter aufnimmt und an das verarbeitende System (SPS, DCS oder Prozessleitsystem) weiterleitet. Vielfach findet sich im Produktionsumfeld nicht nur ein einzelnes System, welches Daten erzeugt. Neben den bereits erwähnten Steuerungssystemen können es auch Labor- oder Inline-Testsysteme sein, die Qualitätsparameter zur Verfügung stellen. Die Gebäudeleittechnik hält Umgebungsdaten wie Raumtemperatur, Luftfeuchte und anderes bereit. Aus der Warenwirtschaft kommen Anforderungsspezifikationen für den Produktionsprozess und herzustellende Losgrößen. Aus der Mittelspannungstechnik und vom Energieversorger kommen Energiemengen, die im Rahmen der Produktion anfallen. Diese sind natürlich nicht auf elektrische Energie beschränkt, sondern können auch Mengen technischer Gase, Dampf, Kälte oder Druckluft sein.

Die Vielfalt der verschiedenen Datenquellen weckt den Wunsch nach möglichst einheitlichen Datenschnittstellen, der jedoch selten bis nie erfüllt wird. So heterogen die Datenquellen sind, so unterschiedlich sind auch die zu betrachtenden Datenschnittstellen. Insbesondere in gewachsenen Produktionsstrukturen beschränken sich die Lieferanten meistens nur auf ihre Komponenten. Umso wichtiger bei der Auswahl einer geeigneten Datenanalyse ist die Vielfalt der angebotenen Schnittstellen, auch über Betriebssystemgrenzen hinweg.

Zu den automatisch erfassten Daten kommen noch die Werte, die über keine technische Anbindung verfügen. Das sind zumeist Handwerte die vor Ort abgelesen werden, wie Wasseruhren und Stromzähler, oder aber Betriebsstundenzähler von Maschinen und Aggregaten aus der Zeit von Industrie 2.0. Es können aber auch Analysedaten aus einem Labor sein, die entweder nur manuell erfasst oder in einem gesammelten Datenformat wie CSV oder XML mit mehreren Parameter in einem Datensatz übertragen werden. Dieser kann auch wieder verschiedene Zeitstempel der jeweiligen Parameter enthalten.

4. Datenspeicher und Aggregation

Sobald die Daten erfasst werden können und die Schnittstellen definiert sind, ist die Datenspeicherung der nächste wichtige Schritt. Im Wesentlichen handelt es sich bei Produktionsdaten um Werte in einer zeitlichen Abfolge (Time Series Data), die miteinander in Korrelation gebracht werden müssen. Zum Beispiel werden Temperaturen aus Reaktoren in gleichen Produktionsphasen miteinander verglichen oder Füllstände in Lagertanks über verschiedene Produktionszeiträume dargestellt. Klassische relationale Datenbanken stoßen hier häufig an ihre Grenzen. Besser sind dort Datenbanken, die auf Time Series Data spezialisiert sind.

Häufig ist aber auch die Manipulationssicherheit der Daten ein wichtiger Faktor. Somit dürfen Daten, die einmal elektronisch erfasst worden sind, nicht mehr verändert werden. Sollte es sich bei den aufgezeichneten Daten aber um einen technisch bedingten Messfehler handeln, kommt man um eine Berichtigung der Werte nicht herum, um im weiteren Verlauf die richtigen Kennzahlen zu erhalten. Dazu ist eine Ergänzung des Wertes um korrigierte Werte unter Beibehaltung des Originalwertes notwendig. Idealerweise kann dieser geänderte Wert auch mit einem Kommentar versehen werden, der eine revisionssichere Auswertung des Änderungsvorganges ermöglicht. Damit ist das gleiche Ergebnis jederzeit überprüfbar. In besonders stark

regulierten Prozessen, wie zum Beispiel der pharmazeutischen Industrie, ist dazu ein Vier-Augen-Prinzip notwendig, damit eine Änderung immer auch von einer zweiten unabhängigen Stelle bestätigt wird.

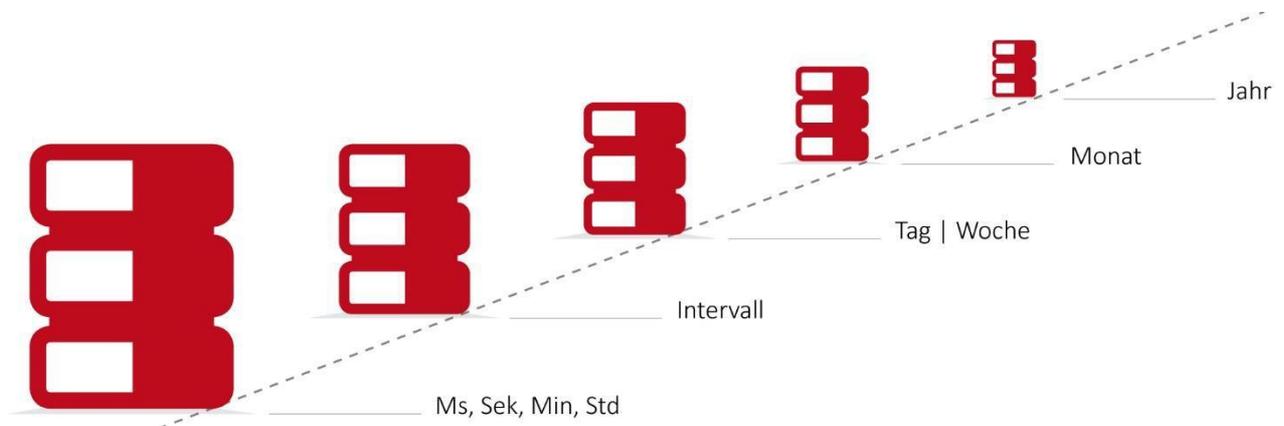


Abbildung 1 Aggregationen

Die so gespeicherten Werte liegen nun auf dem Datenspeicher. Das kann eine Festplatte, eine SSD oder ein NAS System sein, wobei NAS Systeme dabei häufig nicht die beste Wahl sind, da bei Zeitreihenspeicherung ein extrem schneller Zugriff auf den Speicher notwendig ist. Werden im späteren Verlauf viele Daten über lange Zeiträume abgerufen, so entsteht schnell eine große Datenmenge. Daher ist es notwendig die Daten nach bestimmten Rechenvorschriften zu aggregieren, was auch als Verdichten bezeichnet wird. Das Verdichten steht in diesem Zusammenhang nicht für Speicherkomprimierung, sondern für die Einsortierung der Werte in verschiedene aggregierte Datentöpfe. Zum Beispiel: Die Temperatur eines Behälters soll alle 15 Sekunden gemessen und auf den Datenspeicher geschrieben werden. Wird diese Temperatur jetzt über ein Jahr ausgewertet, so müssen alleine für diese Variable 2.102.400 Werte herangezogen werden (4 Messungen pro Minute x 60 Minuten x 24 Stunden x 365 Tage (Schaltjahre nicht mitgerechnet)). Werden 5 Temperaturen, die in diesem Zeitraster aufgezeichnet wurden, in einem Trend über ein Jahr dargestellt, sind das mehr als 10 Millionen Datenpunkte, die grafisch verarbeitet werden müssen. Daher ist die Aggregation mittels mathematischer Berechnungen ein gangbarer Weg. Im ersten Schritt werden die Rohwerte z.B. in mehreren Intervallstufen als arithmetischer Mittelwert verdichtet. Es kann aber auch ein gewichteter Mittelwert, ein Zentralwert (Median), eine Letztwert-Betrachtung (gerne für Energiezähler) oder eine Percentilwert-Berechnung sein. Aus den Intervallwerten werden dann Tages-, Wochen-, Monats- und Jahreswerte gebildet. Der Anwendungsfall sollte hier die Mathematik bestimmen und nicht umgekehrt. Wichtig ist die Flexibilität im Auswertungstool, welche möglichst viele Freiheitsgrade bieten muss.

Die so verdichteten Daten sind eine ideale Grundlage für die späteren Auswertungen. Auch sollten die Rohdaten immer unangetastet bleiben. Wird im späteren Verlauf die Berechnungsvorschrift für die Aggregationsstufen geändert, kann auf Basis der vorhandenen Rohwerte jederzeit eine neue Aggregationsvorschrift verwendet werden.

4.1 Texte

Neben den numerischen Betrachtungen sind auch Texte eine wichtige Kenngröße für eine Datenanalyse. Sie können, wie schon erwähnt, Kommentare zu Wertänderungen enthalten, aber auch eigene Felder sein, die für viele Anwendungsfälle herangezogen werden können. So sind Beispiele von alphanumerischen Chargennummern genauso denkbar wie Hinweise eines Schichtbuches und vieles mehr. Einzig die Berechnungsvorschriften können hier logischerweise entfallen.

5. Berechnungen und Kennzahlen

Werte in den Aggregationsstufen zu berechnen ist aber nur die halbe Wahrheit. Gerade in der Analyse ist auch die Korrelation der Daten untereinander unwahrscheinlich wichtig. Verschiedenste Kennzahlen müssen anwendungsbezogen erstellt und mathematisch wie logisch in den Kontext gebracht werden. Der Energieverbrauch pro Tonne Produkt ist nur ein einfaches Beispiel. Gerade dann, wenn Kennzahlen immer wieder gleich angewendet werden müssen, ist die Verwendung einer Kennzahlenvorlage ein gutes Hilfsmittel, um die Berechnung (z.B. für mehrere Behälter) in gleicher Weise zu betrachten.

6. Auswertung

Um die Daten sicher und übersichtlich auswerten zu können ist eine qualifizierte Datenquelle Pflicht. So ist sichergestellt, dass es nur eine Wahrheit gibt und Entscheidungen auf den realen Produktionsdaten getroffen werden. Die Daten werden entweder im gleichen System gehalten in dem auch die Auswertungen passieren oder es gibt eine gesicherte, performante Schnittstelle zum Datenspeicher, um die Auswertung auf den vorhandenen Datenquellen auszuführen.

Dabei sollen die Daten übersichtlich und vom Anwender situationsbezogen ausgewertet werden können. Grundsätzlich sollten die Datenpunkte strukturiert und organisiert vorliegen, so dass eine einfache Zuordnung des Datenpunktes zu einer Organisationsstruktur oder Ortskennzeichnung möglich ist. Die Strukturen müssen sich aber nicht an festen Ordnern oder Baumstrukturen festlegen. Vielmehr ist es wünschenswert, dass Datenpunkte anwendungsbezogen in verschiedenen Sichten zur Verfügung stehen und dass der Anwender selbst entscheiden kann, wie die Sicht auf die Datenpunkte aussieht. So betrachtet z.B. ein Wartungsmitarbeiter die Datenpunkte und Kennzahlen nach Örtlichkeiten und ein Qualitätsmitarbeiter nach Herstellungsschritten. Die Flexibilität der Darstellung ist hier der Maßstab, den es zu erfüllen gilt, denn auch hier gibt es nicht eine Wahrheit, sondern immer anwendungsbezogene Aufgabenstellungen.

1. Darstellungen

Aus Anwendersicht ist eine Auswertung der Daten auf verschiedenen Endgeräten notwendig. Die Darstellung wird dabei von der Anwendung getrieben. Es ist für die Auswertung eher untergeordnet, ob der Anwender einen Webbrowser oder eine Desktop-Applikation verwendet. Wichtiger ist die einfache Zusammenstellung der Daten gemäß dem Anwendungsfall sowie der Zugriff auf die richtige Zeitbasis bzw. die Korrelation eines oder mehrerer Datenpunkte über verschiedene Zeitachsen in entsprechenden Wertebereichen.

Auch können verschiedene Detailgrade der Datenpunkte wichtig werden. So ist für einen Anwender, der einen Prozess an einer Fertigungslinie begutachten muss, ein hochauflösender Datenstrom eher wichtig als für einen Mitarbeiter aus dem Management, der eine Sicht auf aggregierte Kennzahlen benötigt.

2. Trends und Charts

Die grafische Darstellung in Trends ist wichtiger als bloße Datenkolonnen in Tabellen, da mehr Informationen im Kontext zueinander auf einen Blick betrachtet werden können. Auch hier sind verschiedenste Darstellungen wichtig: Ob zeitliche Verteilung in Linien, Balken oder Zeitpunkte in Torten, Knotenpunkt oder in einer Rasterdarstellung ist wiederum dem Anwendungsfall geschuldet. Aber nicht nur zeitliche Verläufe sind wichtig. In bestimmten Anwendungsfällen können auf der X-Achse auch Datenpunkte aufgetragen werden, die in Korrelation zu den Werten auf der Y-Achse stehen, wie es z.B. bei Ortskurven notwendig ist.

3. Dashboards

Dashboards werden dazu verwendet, verschiedenste Informationen für den Betrachter einfach zugänglich zu machen. Sie stellen die Daten immer in einen bestimmten Kontext. Das können Qualitätsparameter, Energiekennzahlen oder Auslastungswerte sein. Neben der Art der Darstellung ist hier die Flexibilität besonders wichtig. Dem Anwender müssen die Daten in einer für ihn verständlichen Art so aufbereitet werden, dass die daraus resultierenden Informationen leicht ablesbar sind. Selbstverständlich sind hierfür verschiedenste Graph- und Trenddarstellungen wie Linien-, Balken-, Punkt-, Flächen- und Tortendiagramme notwendig. Aber auch einfache Wertanzeigen, Wertetabellen und Ampelsichten der bereitgestellten Daten sind von immenser Aussagekraft. Gepaart mit den unterschiedlichen Datenherkünften wird so aus dem Dashboard eine ultimative Informationsschaltzentrale.



Abbildung 2: Dashboard

Die Erstellung dieser Dashboards darf aber dem Anwender aber keine Programmierkenntnisse abverlangen. Viel wichtiger ist, dass der Anwender, der seine Daten aus Erfahrung kennt, die Übersicht so erstellen kann, dass die spezifische Auswertung seinen Anforderungen genügt und nicht eine im System feste Vorgabe ist. Die Zuordnung verschiedener Zeitbereiche oder Zeitpunkte versteht sich genauso wie die automatische Aktualisierung der Daten, z.B. auf in einer Werkhalle aufgestellten Displays oder in

regelmäßigen Managementmeetings, um aus diesen Daten Verbesserungspotentiale abzuleiten und Erfolge des Unternehmens zu sichern und auszubauen.

6.4 Berichte

Auch die Weitergabe der Daten in einer strukturierten Form als Bericht ist immer noch wichtig. Ob der Bericht dann ausgedruckt wird oder als lesbares Dokument (z.B. PDF) per E-Mail zugestellt wird, ist dabei nicht entscheidend. Wichtig ist eher, dass eine feste Vorlage definiert werden kann, die dann wiederkehrend mit den jeweiligen Daten, z.B. als Tages-, Wochen-, Monats- oder Jahresbericht, befüllt wird. Daten müssen aber

auch über diese starren Grenzen hinweg strukturiert werden können, z.B. in variablen Zeitbereichen oder zu spezifischen Ereignissen wie bestimmten Produktionschargen zusammengefasst. Die Darstellung erfolgt dann wahlweise in Tabellen, einzelnen Kennzahlen- oder Textfeldern sowie Trends.

6.5 Alarme und Ereignisse

Die Betrachtung von Alarminformationen darf im Kontext der Datenanalyse nicht fehlen. Insbesondere die Korrelation von Störungen mit den einhergehenden Produktionsdaten ist ein wichtiges Maß für die Analyse der Produktionsvorgänge. Allerdings ist dieses Thema sehr vielschichtig, so dass an dieser Stelle nur der Hinweis auf die Betrachtung der Kombination von Alarmen und Prozessdaten wichtig ist. Eine umfangreiche Betrachtung zu Alarmanalysen und Eskalationen sind im White Paper "Alarme und Stillstandsmanagement" zu finden.

7. Verfügbarkeit und Datensicherung

Neben der Datenintegrität ist auch die Verfügbarkeit ein wichtiger Faktor bei der Konzeption dieser Informationssysteme. Dies beginnt schon auf der Erfassungsebene. Ist die Datenquelle beispielsweise ein bereits redundant ausgelegtes SCADA System, muss dies auch bei der Erfassung berücksichtigt werden. Die Datenerfassungsfunktion direkt am System muss in dem Fall wissen, ob die Daten vom Master- oder vom Backup-System kommen. Entsprechend des Redundanzstatus muss dann eine Zuordnung passieren, welche Daten verwendet werden, damit keine Dopplungen und somit Verfälschungen in den Rechenvorschriften entstehen.

Wenn die so ermittelten Daten an den Datenspeicher weitergegeben werden und dieser aus verschiedensten Gründen vielleicht nicht verfügbar ist, sollte das Erfassungssystem einen Mechanismus vorsehen, um die Daten so lange zwischenzuspeichern bis der Datenspeicher wieder zur Verfügung ist.

Der Datenspeicher selbst ist nun in der Lage, die Daten an seine Auswertungsbereiche weiterzugeben. Aber auch hier spielt die Verfügbarkeit eine Rolle: Wenn die Auswertung der Qualitäts- oder Energiedaten kontinuierlich Anwendung findet, ist ein Ausfall des Informationssystems nicht tragbar und es muss eine Redundanzfunktion zum Einsatz kommen.

Dass eine Datensicherung Bestandteil eines Prozessdatenspeichers sein muss, ist selbstverständlich. Allerdings liegen auch hier die Unterschiede in der Ausgestaltung. So muss eine Datensicherung ohne Einfluss auf die aktuelle Aufzeichnung genauso möglich sein wie die einfache Rücksicherung der Daten oder einzelner Datenzeiträume in das Produktivsystem.

8. Regulierte Anwendungsfälle

Eine besondere Betrachtungsweise kommt immer dann zum Tragen, wenn die Daten manipulationssicher vorliegen müssen. Hierzu gibt es je nach Branche verschiedene zu erfüllende Standards. In der Pharmaherstellung hat sich der FDA Standard 21 CFR Part 11 etabliert. Die deutsche Abwasserwirtschaft geht

zumeist nach DWA M260 vor, wobei hier einzelne Verbände z.T. eigene Abbildungen und Verschärfungen haben.

Entscheidend ist der bereits in Punkt 3 beschriebene Mechanismus, dass Daten nur dann behandelt werden dürfen, wenn 1.) der Originalwert nicht verändert wird und 2.) eine Kennzeichnung des geänderten Wertes in den Auswertungen möglich ist. Erweitert wird die Änderung dann noch um das Vier-Augen-Prinzip, bei dem zwei voneinander unabhängige Prozessteilnehmer die Wertänderung mit ihrem Kennwort bestätigen.

9. Fazit

Dieses White Paper ist aus mehr als 25 Jahren Erfahrung in diesem Themenbereich entstanden. Von Beginn an steht für VIDEK das Thema der Datenerfassung, -aufbereitung und -bereitstellung ganz oben auf der Agenda. Über die Jahre haben wir uns mit unseren Kunden in tausenden von Applikationen aus diversen Branchen diesem Schwerpunkt von verschiedenen Seiten genähert und dabei gelernt, wo die größten Herausforderungen liegen. Mit ACRON haben wir das führende System im Bereich der Prozessdatenarchivierung im Angebot, mit JUNE5 ein Webinterface auf verschiedenste Datenquellen - so bieten wir in allen erdenklichen Systemarchitekturen eine für Sie optimale Lösung. Gerne erarbeiten wir mit Ihnen Ihren persönlichen Anwendungsfall. Kombiniert mit Kenntnissen zu Cybersicherheitsanforderungen und Regulierungen sind wir mit unserem mehrere hundert Systemintegratoren umfassenden weltweiten Netzwerk Ihr erster Ansprechpartner, wenn es um Historian und Berichte geht. Ihre Impulse sind für uns stetig Antrieb für kommende Innovationen, so dass wir heute und in Zukunft daran arbeiten, aus Ihren industriellen Daten die Informationen zu erstellen, die Sie brauchen, um Ihre Entscheidungen auf der richtigen Basis zu treffen.

Glossar

AGGREGATION

Mathematische Behandlung eines Werteverlaufs zur einfacheren Darstellung z.B. arithmetischer Mittelwert, Median, etc.

SCADA (Scanning Control and Data Acquisition)

Visualisierung von aktuellen Prozesswerten in grafischer Oberfläche. Häufig auch mit HMI (Human Machine Interface) ergänzt.

REDUNDANZ

Gesteigerte Verfügbarkeit durch gleichzeitigen Betrieb zweier paralleler Systeme mit Daten- und Konfigurationsabgleich.

DATENPUNKT

In diesem Kontext immer eine zeitliche Wertereihe mit dahinterliegenden Aggregaten.