



# Von Daten zu Unternehmenswerten

© Sergey Nivens | Adobe Stock

## AUTOREN



**Hendrik Bohlen**  
ist Leiter Technik bei Werum  
Software & Systems AG in  
Lüneburg.



**Dr. Stefan Unterschütz**  
ist Programmverantwortlicher für  
den Bereich Testprozessmanagement  
bei Werum Software &  
Systems AG in Lüneburg.

Der Testprozess in der Fahrzeugentwicklung muss eine hohe Produktqualität bei maximaler Kosteneffizienz und der Wahrung von Compliance-Anforderungen gewährleisten. Allerdings landen Daten aus Simulation und Test allzu oft in Datensilos. Mit einer passenden Data Analytics Toolchain lassen sich die verschiedenartigen Daten verknüpfen und Mehrwerte generieren.

## TOOLCHAIN BINDET SÄMTLICHE TESTDATEN EIN

Die heutige Fahrzeugentwicklung stellt hohe Anforderungen an den Testprozess. Es muss eine hohe Produktqualität bei maximaler Kosteneffizienz gewährleistet werden – unter Erfüllung von Compliance-Anforderungen.

Hierbei fallen wertvolle Daten zum Fahrzeug in Simulationen und Versuchen an. Dazu gehören unter anderem Lebensdauertests, Betriebsfestigkeitsuntersuchungen, Sicherheitsprüfungen und Performancetests. Immer häufiger sind durch den Einsatz von vernetzten, intelligenten Produkten auch Daten aus dem Feld zu erwarten.

Daten werden meist dort gespeichert, wo sie anfallen, in sogenannten Datensilos, und stehen nur selten unternehmensweit zur Verfügung. Dadurch ist es nicht möglich, den Mehrwert aus den Daten zu ziehen, den eine Verknüpfung dieser verschiedenartigen Daten bietet.

Für die digitale Transformation des Testprozesses bietet die moderne IT Lösungen an, um das maximale Potenzial der Daten zu heben. Dieser Beitrag skizziert die Anforderungen an eine Data Analytics Toolchain und eine mögliche Realisierung. **BILD 1** zeigt ein vereinfachtes Prozessschaubild der Toolchain. Es verdeutlicht, über welche Prozessschritte die Daten von der Quelle zum interessierten Anwen-

der beziehungsweise zu weiterführenden Systemen gelangen.

Die Lösung kann unabhängig von bestehenden IT-Systemen aufgebaut werden, und die verschiedenen Datenlieferanten wie zum Beispiel Prüfstände, IIoT-Geräte oder Fachanwendungen können in Stufen eingebunden werden.

## HERAUSFORDERUNGEN BEI DER DATENVERARBEITUNG

Eine moderne Toolchain für den Testprozess muss eine Vielzahl an funktionalen und nichtfunktionalen Anforderungen erfüllen. Zunächst müssen die Testdaten nach der Erfassung vorverarbeitet werden. Datensätze müssen bereinigt werden. Auch ist die Qualifizierung und Anreicherung der Daten um weitere Informationen oder erste statistische Auswertungen erforderlich. In diesem Zuge kann eine Verifikation erfolgen, das heißt die Daten werden auf Vollständigkeit und formale Korrektheit überprüft. Im Rahmen der Vorverarbeitung ist auch eine Harmonisierung und Standardisierung der Daten sinnvoll, was die spätere Nutzung vereinfacht.

Testdaten müssen für die Dauer der geforderten Aufbewahrungsfristen revisionssicher gespeichert werden. Dabei sollten begleitende Informationen, sogenannte Metadaten, zu den eigentlichen Testdaten mit abgelegt werden. Für

eine schnelle Verarbeitung sind kurze Zugriffszeiten wünschenswert. Stark steigende Datenmengen in aktuellen Testfeldern erfordern ein besonderes Augenmerk auf Skalierbarkeit.

Im Umgang mit Daten sind Nachvollziehbarkeit, Transparenz und Auditierung wichtig. Dafür muss dokumentiert werden, wie, auf welchem Prüfstand und mit welchem Sensor Testdaten generiert wurden. Auch jegliche Datenänderung muss protokolliert werden.

Für die Analyse der Testdaten können verschiedene Methoden erforderlich sein. Die Spannweite reicht von Aggregation und Verdichtung, Transformationen, Berechnung von statistischen Kennwerten bis hin zu komplexen Machine-Learning-Algorithmen. Auch die Menge der verfügbaren Analysewerkzeuge variiert stark, was eine hohe Flexibilität bei der Einbindung erfordert.

Die Toolchain muss dem Anwender und externen Anwendungen wie zum Beispiel einem PLM-Werkzeug den Zugriff auf Daten ermöglichen. Damit einher geht der Bedarf nach einer datenübergreifenden Suche und Visualisierung. Eine starke Zugriffskontrolle, häufig in Verbindung mit Freigabeprozessen, ist ebenfalls notwendig. Neben einer auf Rollen und Rechten basierenden Zugriffsteuerung sind oft weiterführende Mechanismen erforderlich, um Daten zu klassifizieren und zu schützen.

Für diese Daten müssen die Unternehmen zunehmend Compliance-Anforderungen erfüllen, zum Beispiel Informationsschutzanforderungen nach DSGVO, ISO/IEC 27001, TISAX oder zum Nachweis der erforderlichen Produkteigenschaften wie zum Beispiel WLTP oder Homologation. Hinzu kommt der Nachweis der Kompetenz von Testlaboren gemäß DIN EN ISO 17025.

Gegebenenfalls sind noch weitere Anforderungen zu erfüllen: Neben Performance sind auch Ausfallsicherheit, Verfügbarkeit, Zuverlässigkeit, Fehler-toleranz und Verteilbarkeit zentrale Anforderungen an alle Dienste der Toolchain. Abschließend sind die Anforderungen für den Betrieb des Systems zu definieren. Wichtige Stichworte sind hier Orchestrierung und Monitoring.

Die genannten Anforderungen müssen vom Unternehmen hinsichtlich ihrer Relevanz bewertet werden, was signifikante Auswirkungen auf das Design der Toolchain hat. Sofort einsatzbereite Lösungen sind zum aktuellen Zeitpunkt nicht verfügbar: zu unterschiedlich sind die konkreten Unternehmensanforderungen, zu wechselhaft die anzuwendenden Regularien und Gesetze, zu vielfältig die Tools, um Daten auszuwerten.

Im Folgenden wird vorgestellt, wie eine moderne Toolchain aufgebaut wird. Soweit möglich und sinnvoll, kommen

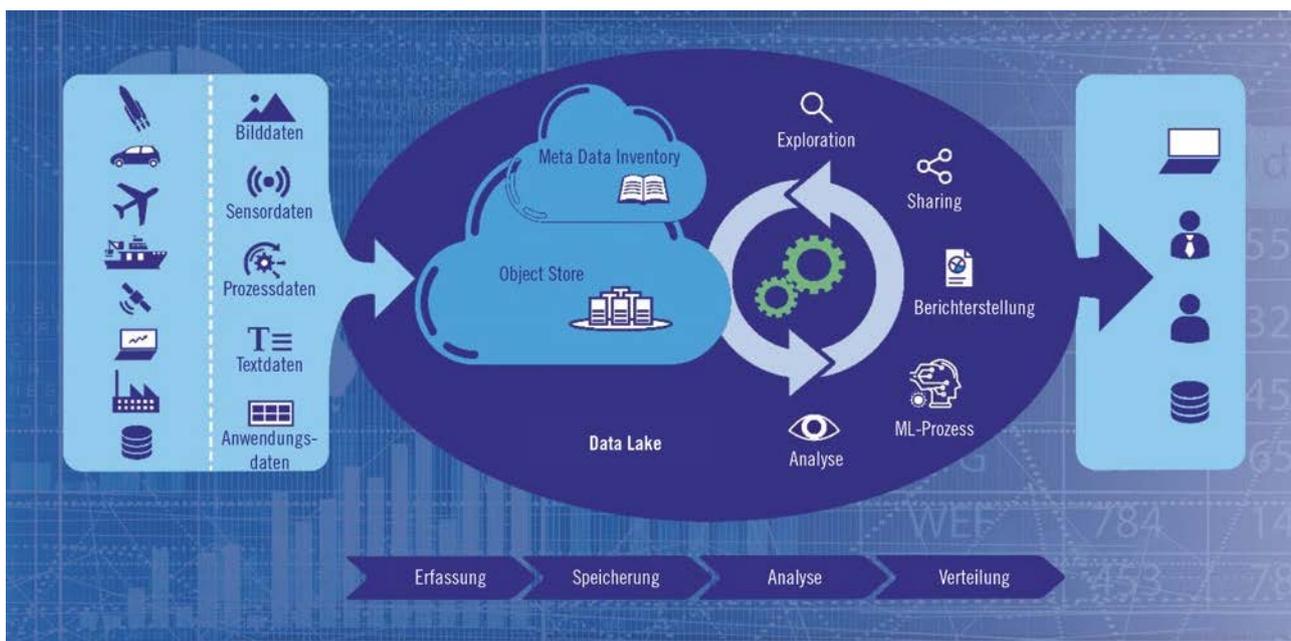


BILD 1 Prozess-Schaubild der Data Analytics Toolchain (© Werum)

die genannten Werkzeuge aus dem Open-Source-Bereich. Bei Bedarf werden aber auch kommerzielle Pendant eingesetzt, wenn diese für bestimmte funktionale Anforderungen notwendig sind.

**VON DATEN UND METADATEN**

Für die Verarbeitung von Daten bietet sich als versinnbildlichte Analogie die Verarbeitung von Rohstoffen an. Rohstoffe werden über komplexe Arbeitsschritte zu immer höherwertigen Produkten weiterveredelt, bis das gewünschte Endprodukt entsteht. Denselben Mechanismus verwendet eine Toolchain. Rohdaten werden in höherwertige Datenprodukte umgewandelt, die ihrerseits in weitere Datenprodukte umgewandelt werden können.

Neben der Transformation und Analyse von Daten spielt die Anreicherung um Metadaten eine zentrale Rolle. Metadaten sind Informationen über Daten. Ohne Metadaten sind Messdaten für eine sinnvolle Weiternutzung im Unternehmen wertlos. Metadaten sind in verschiedene Kategorien unterteilbar:

- Beschreibende Metadaten, die Daten identifizieren, auffindbar machen und mit weiterem Kontext anreichern, zum Beispiel wann was von wem mit welchen Instrumenten gemessen wurde

- strukturelle Metadaten, die Auskunft über Form und Aufbau der Daten geben
- administrative Metadaten, die es ermöglichen, Daten zu verwalten
  - zum Beispiel Zugriffsrechte
- statistische Metadaten, die statistische Auskunft über Daten geben
  - zum Beispiel KPIs.

Metadaten sind für die Verknüpfung und Verarbeitung von Daten in einer Toolchain von zentraler Bedeutung. Existierende Fachanwendungen, die Daten produzieren, müssen gegebenenfalls erweitert werden, damit notwendige Metadaten zur Verfügung stehen.

**AUFBAU DER DATA ANALYTICS TOOLCHAIN**

**BILD 2** zeigt den funktionsorientierten Aufbau der Toolchain. Die Toolchain ist zwischen den Produzenten und Konsumenten von Daten angesiedelt. Typische Produzenten sind Prüfstände beziehungsweise Mess-einrichtungen oder zukünftig beliebige IIoT-Geräte. Daten können aber auch manuell erfasste Kennwerte oder Fotos sein, die im Zuge einer Befundung aufgenommen worden sind. Neben den Messdaten selbst kommen als Metadaten die Beschreibung des Testauftrags und des Versuchs sowie mögliche Kommentare und vom Prüfer dokumentierte Vorkommnisse hinzu.

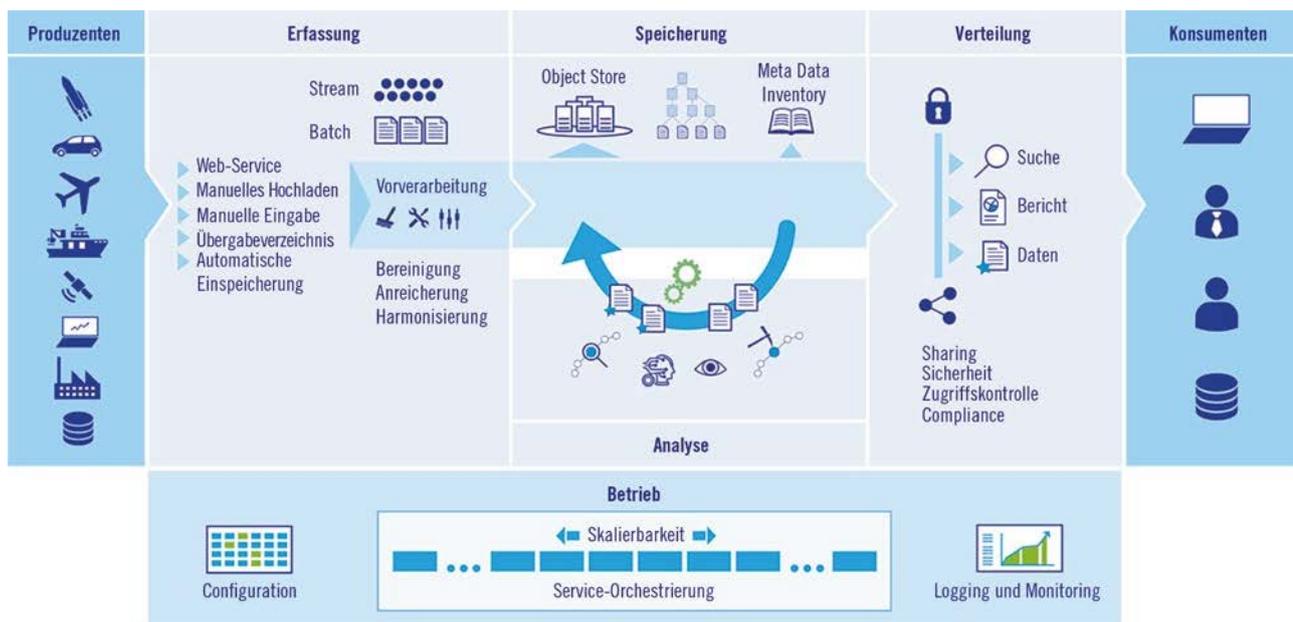
Nachfolgend wird die Funktionalität beschrieben, die durch entsprechende Dienste und Fachanwendungen umgesetzt wird.

**VOM PRODUZENTEN ZUM KONSUMENTEN**

Der Upload erfolgt manuell, zum Beispiel durch Kopieren von Dateien, über ein Web-Frontend, als Eingabe von Einzelwerten, oder automatisiert, zum Beispiel über eine REST-API oder ein Übergabeverzeichnis. Der nächste Schritt ist die Datenaufnahme und Vorverarbeitung. Er kann als Batch- oder Stream-Prozess umgesetzt werden. Letzteres ermöglicht eine zeitnahe Datenanalyse.

Wesentliche Funktionen der Vorverarbeitung sind die Bereinigung und Anreicherung sowie Harmonisierung und Standardisierung der Daten. Weiterhin werden die Metadaten so aufbereitet, dass eine Suche und Weiterverarbeitung möglich wird. Zum Abschluss werden Daten und Metadaten in einem Store abgelegt.

Für die Speicherung der Daten bietet sich ein Object-Store an. Er ist für die Speicherung von großen Mengen unstrukturierter Daten optimiert und hat sich in den letzten Jahren gegen dateibasierte Speicherung oder auch Apache HDFS durchgesetzt. Object-



**BILD 2** Funktionsorientierter Aufbau der Toolchain (© Werum)

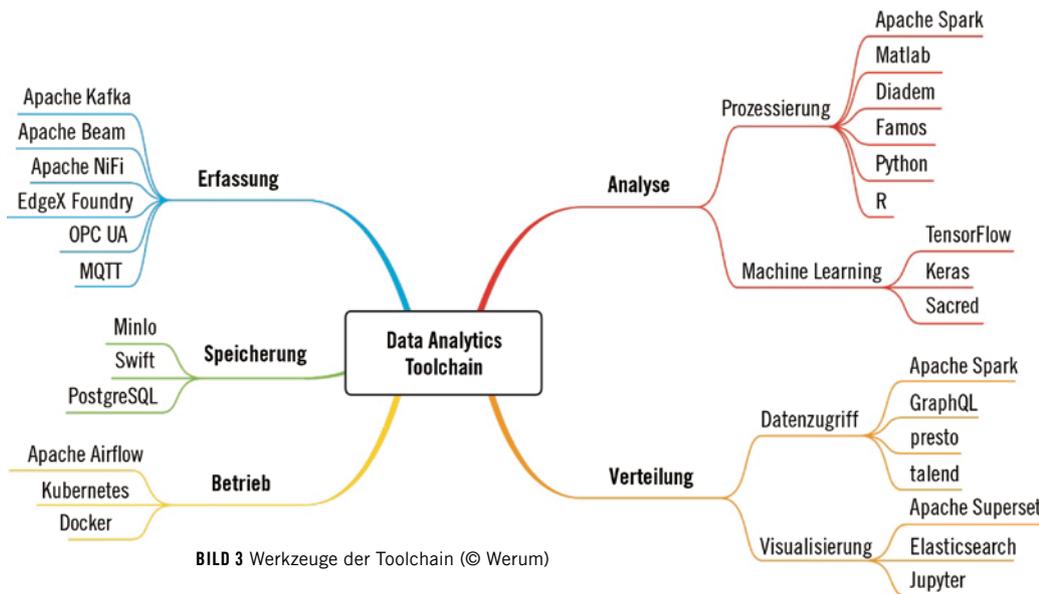


BILD 3 Werkzeuge der Toolchain (© Werum)

Stores bieten eine nahezu unbegrenzte Skalierbarkeit bei geringen Kosten und kurzen Zugriffszeiten.

Neben den Daten selbst können Object-Stores auch deren Metadaten speichern. In der Praxis hat sich jedoch der Einsatz einer relationalen Datenbanklösung als vorteilhaft gezeigt, die Metadaten in strukturierter Form sowie eine Referenz auf die Daten selbst speichert, ein sogenanntes (Meta-)Data-Inventory. So können Metadaten leichter angepasst und ausdrucksstärkere Suchanfragen definiert werden und die Datenprozessierung kann mithilfe der Metadaten leichter umgesetzt werden.

Die IT-Infrastruktur muss die Verteilung der Daten gewährleisten, damit die Konsumenten auf die Daten und Metadaten zugreifen können. Zugriff beziehungsweise Zugriffskontrolle sind dabei mehrdimensional zu betrachten. Neben einem rollenbasierten Modell kann der Zugriff auf Daten auch über Metadaten wie zum Beispiel Standort oder Freigabestatus gesteuert werden. Für diese komplexen Anforderungen müssen meist spezialisierte Lösungen hinzugezogen werden.

Während der Datenverteilung kann eine Aufbereitung oder Transformation erfolgen. Als Ergebnis können benötigte Datenformate oder ganze Reports zur Verfügung gestellt werden.

## ANALYSE UND AUSWERTUNG

Die bisherige Funktionalität der Toolchain ist Grundvoraussetzung für die Datenanalyse und die Verarbeitung zu höherwertigen Produkten.

Neben der explorativen Analyse und Auswertung einzelner Daten mit entsprechenden Werkzeugen können auch standardisierte Auswertungsroutinen eingesetzt werden. So kann ein Auftragsworkflow etabliert werden, der Testaufträge von einem PLM-System entgegennimmt. Die Testdaten werden anschließend den Testaufträgen zugewiesen und können prozessgesteuert verarbeitet werden. Finales Ergebnis kann der Bericht für einen Auftraggeber sein, der vorab vom Auftragnehmer freigegeben wird. Der Vorteil der Toolchain ist es, den Werkzeugen alle angefallenen Daten über eine standardisierte Schnittstelle zur Verfügung stellen zu können und dabei eine Zugriffskontrolle zu gewährleisten.

Für die skalierbare und parallelisierte Prozessierung von vielen Daten bietet sich Spark an. Spark ist Quasistandard und früheren Ansätzen wie Apache Hadoop überlegen.

Eine weitergehende Nutzung der Daten kann mittels Machine Learning erfolgen und komplett neue Einblicke und Anwendungsfelder eröffnen. Ansätze wie Predictive Maintenance können in Dauerlauf- und Feldversuchen eingesetzt werden, um frühzeitig Anomalien oder mögliche Ausfälle von Komponenten zu identifizieren. Machine Learning kann auch genutzt werden, um den Prüfer auf „verdächtige“ Datensätze hinzuweisen. Das ist umso wichtiger, da zukünftig mit einer ansteigenden Menge an Daten in Versuchen zu rechnen ist, sodass eine maschinell unterstützte Datenbewertung unabdingbar wird.

## BETRIEB UND WERKZEUGE

Die Toolchain kann als dienstorientierte Architektur aufgebaut werden. Für die Orchestrierung, Konfiguration und Überwachung aller Dienste werden weitere Werkzeuge benötigt. Mithilfe einer Workflow Engine können zusätzlich die oben genannten Prozessierungsschritte verwaltet und automatisiert ausgeführt werden.

Die Entscheidung, die Toolchain On-Premise, als Corporate Cloud oder als Public Cloud zu betreiben, hängt stark von den jeweiligen Unternehmensanforderungen ab. Neben Kosten stehen IT- und Compliance-Anforderungen im Vordergrund. BILD 3 zeigt mögliche Werkzeuge für den Aufbau der Toolchain.

## FAZIT

Um das Potenzial der Daten zu heben und Datensilos aufzubrechen, ist eine Data Analytics Toolchain von zentraler Bedeutung. Der Überblick über die vorhandenen Werkzeuge und ihr Zusammenspiel zeigt, dass die heutigen Anforderungen an den Testprozess mit einer passenden Toolchain lösbar sind. Ihr Einsatz kann die Transformation im Bereich des Testprozesses erfolgreich vorantreiben.

Die zur Verfügung stehenden Werkzeuge müssen dann noch in das Unternehmensumfeld integriert werden, denn es gibt es keine passende Toolchain, die sofort einsatzbereit ist und die Anforderungen des Unternehmens erfüllt.

