

Foto: Gerd Altmann / Pixabay



Trends und Entwicklungen auf dem Data & Analytics-Markt

Der Markt für Data & Analytics befindet sich in einem gewaltigen Umbruch. Neue Speichertechnologien und entsprechende Frontends bedeuten eine echte technische Revolution, die sich bereits in der Unternehmenspraxis niederschlägt. QUNIS fasst die Entwicklungen am Softwaremarkt und aus aktuellen Kundenprojekten zusammen und gibt einen Überblick über Trends, die bei der Konzeption und Erweiterung von datengetriebenen Applikationen zu beachten sind.

ORGANISATIONEN forcieren derzeit Digitalisierungsinitiativen auf allen Ebenen. QUNIS verzeichnet in den aktuellen Projekten einen enormen Innovationsschub mit sehr

breit gestreutem Anwendungsspektrum. Alle Initiativen haben eine Gemeinsamkeit: Es werden riesige Datenräume eröffnet, die ganz neue Anforderungen an die Speicherung

und Verarbeitung von Daten stellen. Der Bereich Data & Analytics muss sich diesen Anforderungen stellen. Die Ausweitung der Themen und Aufgaben in der Analytics-Welt erfordert

technologisch und organisatorisch neue Ansätze.

Revolution in der Datenbank-Entwicklung

Eine echte Revolution findet dabei auf der technologischen Seite statt. Wir bewerten die derzeitigen Entwicklungen im Umfeld von Data & Analytics als die innovativste Phase der letzten 30 Jahre. Entscheidend ist dabei die Entwicklung der Datenbanken. Über Jahrzehnte waren im BI- und Analytics-Markt die relationalen und multidimensionalen Datenmodelle Standard, die sich optimal für die Ablage und Verarbeitung strukturierter Daten eignen. In diesen Modellen laufen Storage und Computing gemeinsam in einer Umgebung und sind damit auch an die Limits der jeweiligen Technologie gebunden. In der klassischen Data-Warehouse-Architektur mit dem SQL-Server beispielsweise bestimmt die SQL-Datenbank die Skalierbarkeit, die Art der Abfragen und die Entwicklungs- und Abfragesprache. Diese Grenzen werden nun durch die Trennung von Storage und Compute gesprengt!

Massiv parallele Datenspeicherung

Neben den klassischen Datenbanken kommen zunehmend mächtige Storage-Technologien als zentrale Datenspeicher für die gesamte Organisation zum Einsatz. Hier setzt sich die massiv parallele Datenspeicherung durch. Diese Datenbanken bringen keine Compute-Funktionalität mit, sondern sind ausschließlich auf die Speicherung umfangreicher und heterogener Datenmengen spezialisiert. Die Verarbeitung, das Processing und die Aufbereitung von Daten, laufen in diesem Architekturmodell getrennt vom Speicher in eigenen hochspezialisierten Prozessoren.

Vorteile der neuen Datenspeicher:

- **Hohe Performance:** Durch die massiv parallele Skalierung über Cluster-Verbünde können Daten

sehr schnell gespeichert und abgefragt werden.

- **Heterogene Datenwelten:** Es können unterschiedlichste Datenformate abgelegt werden. So lassen sich beispielsweise JSON-Datensätze aus der Produktion mit Tabellenformaten aus einem ERP-System und Textelementen aus dem Internet oder dem Customer Support in ein und derselben Technologie speichern.

Flexibilität durch Trennung von Storage und Compute

Massiv parallele Speicher eignen sich damit optimal als Shared Service für unterschiedlichste Applikationen, so auch als zentrale Grundlage für die breit gestreuten Anforderungen der Analytics-Welt. Durch die separaten Compute-Technologien, die auf den Datenspeicher zugreifen, erhöht sich die Flexibilität im Anwendungsbereich nochmals erheblich. Es lassen sich ganz neue Szenarien einrichten. Unterschiedlichste Use Cases können dabei auf einer gemeinsamen Datenbasis abgebildet werden. Wer wie viel Compute für welchen Use Case auf dem gemeinsamen Storage bekommt, kann in dieser Architektur detailliert gesteuert werden.

Die separierte Storage- und Compute-Architektur wird beispielsweise für Anwendungen wie Change Data Capture, für Delta-Ermittlungen und für die Aufbereitung relationaler Strukturen eingesetzt. Weitere Anwendungsmöglichkeiten sind das hoch performante Training von ML-Cases auf dem Datenspeicher, die Exploration von Rohdaten oder der Aufbau von Schnittstellen für die Vernetzung von Maschinen.

Anwendungsbezogene Konzeption der Plattform

Durch die Trennung von Storage und Compute kann also ein wesentlich breiteres Anforderungsspektrum abgedeckt werden. Wegen der Vielfalt der Möglichkeiten, die moderne Architekturen bieten, spricht QUNIS inzwischen von Datenplattformen.

Mehr Möglichkeiten heißt allerdings auch, dass bei der Architekturkonzeption mehr zu bedenken und zu entscheiden ist. Welcher Ansatz für welche Organisation der richtige ist, hängt allein von den geplanten Anwendungen ab. Eine Applikation für Reporting und Visualisierung beispielsweise ist heute für jede Organisation selbstverständlich. Aber wie weit will ein Unternehmen dabei Self-Service BI ausprägen? Wird ein Data Science Workplace benötigt? Ist eine Streaming-Anwendung geplant und sind API-Strategien nötig oder müssen keine Daten weitergegeben werden? Diese und weitere Fragen bestimmen die Konzeption der Datenplattform.

Lösungsorientierte Implementierung von Kernarchitekturen

Die IT-Architektur bildet den Kern der Datenverarbeitung und -speicherung in Data & Analytics-Initiativen. QUNIS hat Lösungspakete für fünf Kernarchitekturen entwickelt: **Self-Service-BI, Data Warehouse, Data Warehouse & Data Lake, Lakehouse** und **Streaming**. Jede Datamanagement-Architektur enthält folgende Komponenten: Compute für Transformation, Speicher für strukturierte Daten, Governance & Security, Daten Integration & Ingestion sowie Serving Layer für das Frontend.

Je nach Anforderungen kann die vorhandene Systemarchitektur beim Kunden damit zielgerichtet und auf der Basis von Best Practices ergänzt und zur unternehmensindividuellen Datenplattform ausgebaut werden. Das lösungsorientierte Vorgehen bringt Ordnung in den derzeit überbordenden Technologiemarkt und gewährleistet ein strukturiertes Vorgehen mit sicheren Projekterfolgen.

Was tut sich in der Organisation?

Die Organisation rund um die Daten hat den gleichen Stellenwert wie die Technologie. Hier ist derzeit in den Unternehmen einiges im Umbruch.

Vor allem schreitet die „**Demokratisierung der Daten**“ weiter voran. In einer mittleren bis großen Organisation können nicht mehr alle Datenprozesse von einem einzigen zentralen Bereich gestemmt werden. Fragen wie „Wie setzt man Daten im Prozess ein, wie generiert man Insights daraus, wie werden Kennzahlen definiert und Daten aufbereitet?“ erfordern Know-how von allen Fachbereichen.

Auch die Datenqualität kann nur im Rahmen unternehmensweiter Initiativen mit den jeweiligen Data Owners gesichert werden. Im gesamten Unternehmen sind Datenkompetenzträger gefragt. Tatsächlich ist bei vielen Kunden derzeit zu beobachten, dass vorhandene zentralisierte BI- oder Digitalisierungs-Organisationen aufgebrochen werden, um sie breiter im Unternehmen anzulegen. „**Data-driven**“ erfordert eine **Enterprise-Datenorganisation!**

Moderne **Frontends** unterstützen die Verlagerung von Datenkompetenz in die Fachbereiche. Viele Frontends enthalten inzwischen Datenmanagement-Funktionen, mit denen Power User selbst auf verschiedene Datenräume zugreifen und ihre eigenen Datenmodelle aufbauen können. Die weitgehenden **Self-Service**-Möglichkeiten machen die User von Konsumenten zu Gestaltern ihrer eigenen Analyseprozesse und ermöglichen den Business-Funktionen damit ein sehr agiles Entscheiden und Handeln. Die Kehrseite ist aus der Self-Service-BI bekannt: Es besteht die Gefahr der Förderung von Datensilos und Insellösungen. Die Ausgestaltung von Self-Service-Szenarien mit mächtigen Frontends ist daher unbedingt im Rahmen des Gesamtkonzepts von Enterprise-Architekturen mit zu bedenken.

In diesem Kontext ist auch eine definierte **Data Governance** unverzichtbar. **Stamm- und Metadatenmanagement** sowie ein unternehmensweites **Datenqualitätsmanagement** sind Pflichtdis-

ziplinen, die bei der Gestaltung von Data & Analytics-Plattformen von Anfang an zu berücksichtigen sind.

Schrittweise Umsetzung als Cloud-Lösung

Unser Fazit lautet, dass für die aktuellen Anforderungen **Data & Analytics-Plattformen nötig sind**. Durch die massiv parallelen Storage-Konzepte und die Trennung von Storage und Compute haben sich dabei ganz neue Freiheitsgrade für die Realisierung verschiedener Use Cases aufgetan. Das jahrzehntelang vorherrschende Data Warehouse ist in der Data & Analytics-Welt heute nur noch eines der möglichen Architekturmodelle.

Die **Cloud** eröffnet dabei immer mehr technologische Flexibilität. Auf einer **Cloud-Plattform** können problemlos Hersteller-übergreifende Technologien betrieben werden, wobei auch die Grenzen zu nicht kommerziellen Lizenzen und Open-Source-Technologien inzwischen verfließen. Architekturen können damit sehr zielgerichtet und individuell aufgebaut werden.

In der Praxis werden zunehmend **BI-Architekturen zu Data & Analytics-Plattformen** ausgebaut, um beispielsweise im ersten Schritt **Process Mining** oder **ML** zu etablieren. Ein Hinweis als Mutmacher: Bestehendes Know-how wie Businesslogiken ist bei dieser Vorgehensweise wiederverwendbar. Falls gewünscht, lässt sich die gesamte BI-Umgebung in einer modernen Plattform abbilden und diese mit fortgeschrittenen Mo-



Steffen Vierkorn ist Geschäftsführer der QUNIS GmbH. Seine Schwerpunkte liegen in der Architekturkonzeption von Data & Analytics-Systemen, der Entwicklung von BI-, Advanced-Analytics- und Big-Data-Strategien sowie dem Aufbau adäquater Organisationen. Neben seiner Tätigkeit bei QUNIS lehrt Steffen Vierkorn an der TU München und ist Trainer für die CA controller akademie.

dulen erweitern. Es muss keinesfalls alles auf der grünen Wiese neu konzipiert werden.

Unsere Empfehlung für die Umsetzung lautet: Schritt für Schritt vorgehen. Durch die hohe Flexibilität auch in der Skalierbarkeit, die Hyperscaler heute bieten, werden auch die Planung, die Realisierung, das Deployment und die Anpassung von Enterprise-Plattformen einfacher. Das Volumen an Storage und Processing, Query-Zeiten und Userzahlen sind kurzfristig erweiterbar oder zurückzufahren; diese Kategorien müssen heute nicht mehr auf Jahre im Voraus geplant und gebucht werden. *we*