

„Big Data erschließt neue Informationsquellen“

Interview mit Roy Kronester

OBJEKTSpektrum: Big Data ist eines der wichtigsten IT-Trendthemen der letzten Jahre. Welche praktische Bedeutung hat das Thema aus Ihrer Sicht für die Unternehmen?

Roy Kronester: Wenigen Unternehmen ist bewusst, dass das Konzept zu Big Data mehr umfasst, als nur riesige Datenmengen zu verarbeiten. Vielmehr geht es um einen holistischen Ansatz, Daten zu sammeln, zu verarbeiten, zu analysieren und Informationen aufzubereiten, die die Basis für Geschäftsentscheidungen darstellen. Hinzu kommen extrem schnelle Reaktionszeiten, die Informationen kontinuierlich verfügbar machen und dabei neue Datenquellen, wie z. B. soziale Netzwerke, miteinbeziehen.

Wo stehen die Unternehmen bei der Umsetzung von Big Data?

Viele Unternehmen nutzen bereits Big-Data-Ansätze, setzen das Konzept jedoch noch nicht ganzheitlich um. Eine umfassende Betrachtung ist jedoch essenziell, um komplexe Daten zu strukturieren und in verständlich aufbereiteter Weise auszugeben. Die Implementierung eines solchen Ansatzes bedarf dementsprechend einer ausführlichen Beratung. Zunächst muss eine Bestandsaufnahme durchgeführt werden, um festzustellen, ob die benötigten Ressourcen für eine Big-Data-Lösung gegeben sind. Daraufhin werden die Bedürfnisse und Strategien von Unternehmen und Fachabteilungen priorisiert, es wird ein passendes



Roy Kronester ist Director Data & Analytics und Technical Solution Professional – Business Intelligence bei der Avanade Deutschland GmbH.

Konzept erstellt und letztendlich werden die Ergebnisse ausgerollt.

Dieser Ansatz dürfte denjenigen bekannt vorkommen, die den BI-Hype (Business Intelligence) bereits von Anfang an mitgemacht haben, oder?

Tatsächlich sind die grundlegenden Ansätze von BI und Big Data sehr ähnlich und unterscheiden sich im Wesentlichen nur durch einige neue technologische und strukturelle Ideen. Zum einen können heute, im Gegensatz zu den BI-Anfängen, leistungsfähigere

Systeme den gewaltigen Datenstrom besser und schneller bewältigen. Zum anderen müssen neue Informationsquellen, wie z. B. soziale Netzwerke oder Log- und Prozessdaten, ausgewertet werden, die in der Vergangenheit nur rudimentär oder gar nicht im Fokus der Anwender standen.

Beispiele für den technologischen Fortschritt sind in diesem Zusammenhang die Möglichkeiten der parallelen Verarbeitung durch Systeme, die sich nahtlos in die bestehende Infrastruktur integrieren lassen. So ist es Unternehmen möglich, dem ständig wachsenden Daten- und Informationsstrom auf der prozessualen Seite angemessen zu begegnen. Des Weiteren können über ein solches System auch bestehende Lösungen so optimiert werden, dass sie Daten auswerten, die bisher aus Laufzeitgründen nur selten bis gar nicht betrachtet werden konnten.

Wo sind die strukturellen Unterschiede?

Mit dem Wunsch, immer näher an den Ursprung relevanter Informationen zu gelangen, werden zusätzliche Datenquellen identifiziert, die bisher nicht oder nur sehr schwer auswertbar waren. Egal ob maschinengenerierte Prozessdaten oder Twitter- und Facebook-Posts: Diese Informationen können als signifikante Grundlagen zur strategischen Entscheidungsfindung herangezogen werden. Das bedingt jedoch die Möglichkeit, „semi-strukturierte“ Daten in einer geeigneten Datensinke ablegen zu können. In diesem Zusammenhang spricht man innerhalb der „Big-Data“-Initiative von einer „Polybase“-Architektur. In die-

sem Modell werden wenig strukturierte Informationen in einem Dateisystem abgelegt und bei Bedarf durch eine erweiterte Standardabfrage mittels T-SQL vollkommen transparent mit den vorliegenden Informationen aus den bestehenden zentralen Datenbanken kombiniert.

Aber damit hat man eine Unmenge von unterschiedlichsten Daten gesammelt. Wie bekomme ich diese ausgewertet?

Eine Auswertung dieser Informationen ist neben der technischen Implementierung nur durch eine kreative und fachliche Analyse möglich. Die hierfür notwendigen Kenntnisse müssen in den verantwortlichen Teams meist jedoch noch ausgebaut werden.

Zusätzlich zur Bereitstellung von Informationen gewinnen die notwendigen Werkzeuge zur Auswertung komplexer Datenquellen immer mehr an Bedeutung. Die zentrale IT kann nicht alle aufkommenden Fragen der Anwender in ausreichender Zeit bearbeiten und der Ruf nach einer leicht zu bedienenden Oberfläche wird immer lauter. In diesem Zusammenhang stellt sich MS Excel als ein beliebtes Werkzeug dar. Es hat sich in der Analyse von Informationen seit jeher bewährt und lässt sich einfach um neue Funktionalität ergänzen. In der neuesten Generation von MS Office wurden da-

her die Analysewerkzeuge zur Darstellung und Auswertung von komplexen Datenquellen nochmals erweitert. Der Benutzer ist durch diesen „Self-Service-BI“-Ansatz in der Lage, aus einem vielseitigen Datenangebot einen individuell zusammengestellten Datencocktail zu mixen.

Und wie führe ich Big Data im Unternehmen ein?

Ausgangspunkt für die Entwicklung einer ganzheitlichen Informationsstrategie ist die Zieldefinition. Dabei werden die Möglichkeiten von Big Data abgesteckt. Fragen wie „Welche Daten brauche ich?“, „Wie komme ich an die für mein Unternehmen notwendigen Daten?“ und „Darf ich die benötigten Daten überhaupt nutzen?“ müssen vorab geklärt werden. Die Fachabteilungen sollten in diesem sehr frühen Stadium eingebunden werden, da sie mit den produzierten Ergebnissen arbeiten werden. Die Erstellung eines umfangreichen Anforderungs- und Fragenkatalogs ist somit ein essenzieller Bestandteil eines jeden Big-Data-Projekts.

Auf der technischen Seite stehen die Rechenleistung und der Datenspeicher im Vordergrund. Je nach Volumen des Ansatzes ist dieser Bereich intern auszubauen oder über eine Zusammenarbeit mit einem entsprechenden Dienstleister abzudecken. Eine skalierbare Datenplattform lässt sich mit Cloud-Technologien und Managed-Services-Konzepten auch dann aufbauen,

wenn entsprechende lokale Strukturen fehlen oder noch ausgebaut werden müssen.

Personelle Ressourcen spielen ebenso eine signifikante Rolle, da Fachabteilungen und einzelne Mitarbeiter bei der Umsetzung des Ansatzes eingebunden werden müssen. Neben der technischen Umsetzung ist eine Lösung auch intern entsprechend zu vermarkten. Beides ist wichtig für die Akzeptanz der Big-Data-Lösung durch Mitarbeiter und Fachabteilungen, denn nur dann kann ein langfristiger Erfolg gewährleistet werden.

Bei der abschließenden Entscheidung, ob der komplette Ansatz in einem oder mehreren Schritten erfolgt, sind sowohl finanzielle als auch strategische Aspekte zu betrachten. Diese erste Analyse- und Assessment-Phase – von der Entscheidung über einen holistischen Ansatz bis hin zur Frage der Implementierung – sollte innerhalb weniger Wochen abgeschlossen werden, denn der anschließende Konzeptionszeitraum kann durchaus drei bis sechs Monate in Anspruch nehmen. Das finale Roll-Out wird meist in ein bis zwei Jahren abgewickelt.

Das Interview führte ...

|| ... Thorsten Janning
Chefredakteur OBJEKTSpektrum.