DataNavigator

EUROGROUP CONSULTING

DataNavigator #5

Mai 2021

ISSN 2702-1904

Ja, wo sind sie denn? – Daten im IT-Betrieb

Kurt Tanneberger

Auf den ersten Blick legt die Lektüre der täglichen IT-Meldungen und Newsletters nahe, dass „alle und alles in die Cloud“ gehen sollte und damit „alles agil“ und somit leichter wird. Mittlerweile haben die meisten Unternehmen zumindest initiale Stufen einer Cloud-Migration durchlaufen. Manchen Studien zufolge werden bis 2025 bis zu 80% der Unternehmen ihre IT in die Cloud migriert haben. Wenn man diesem Trend aber folgt, sind nicht zuletzt einige Datenbezogene Herausforderungen zu meistern, bevor man aus der Cloud-Plattform den erwarteten Nutzen ziehen kann. Die Disziplin Data Governance, die in klassischen Umfeldern oft noch unvollständig ist, wird mit der Cloud nicht einfacher - im Gegenteil. Ein wesentlicher Unterschied ist, dass ein Cloudanbieter die Daten und üblichen Wege im Unternehmen nicht kennt - und auch nicht kennen soll - und nicht aus eigenem Antrieb oder guter Gewohnheit die nötigen Arbeiten und Kontrollen durchführt, die niemand so wirklich spezifiziert oder gar wirksam bestellt hat.

Sind damit, wenn man die Perspektive eines sicheren und effizienten IT-Betriebs einnimmt, die Fragen des Datenmanagements aus der „Legacy Welt“ im aktuellen Umfeld noch invariant gültig, fallen sie einfach weg oder verändern sie sich nur - und damit zwangsläufig auch der Umgang damit? Wird der IT-Betrieb speziell im Bereich Datenhaltung und -management durch Verlagerung von IT-Services in die Cloud einfacher und günstiger? Oder ist etwa das Gegenteil der Fall? Worauf ist zu achten?

Dazu werden in dieser Folge des Data Navigator einige wichtige Datenmanagement-Herausforderungen dargestellt und im Weiteren unter dem Blickwinkel einer „Cloud-Brille“ betrachtet.

Fragestellungen bisher in der klassischen Welt

Datenwachstum – Datenspeicherung, -sicherung und -wiederherstellung – Housekeeping

Die Bewältigung ständig wachsender Datenmengen begleitet die IT seit ihren ersten Tagen, sowohl für aktive Daten als auch deren Sicherung oder Archivierung. Archivierung und Housekeeping - also die Auslagerung nicht mehr aktiv benötigter Daten aus der Datenbank mit

nachfolgendem Aufräumen / Löschen im aktiven Bereich sind auch nach wie vor wichtige Aufgaben, wenn man Systeme wie SAP technisch beherrschbar und performant halten will. Von den Dateilaufwerken in den Firmen, deren Inhalt und Berechtigungen oft ein wirtschaftlich nicht mehr bezähmbares Gewirr darstellen.

Das Wachstum kennt kein sichtbares Ende - in den 90er Jahren war ein Gigabyte an Daten noch eine respektable und nicht gerade preiswerte Größenordnung, heute haben viele Firmen einen Bestand im Bereich von Petabytes aufgebaut.

Um Kosten zu sparen, hat man hierarchische Speichersysteme eingerichtet, mit rarem schnellem, aber teurem Speicher für die performance-kritischen Anwendungen (oder die, die sich erfolgreich so darstellten), einem Mittelklassespeicher für normale Anforderungen und langsame, günstige Speicherarten für selten genutzte oder ganz passive Daten. In der guten alten Zeit bewegten die Mainframes die nicht mehr so aktiv genutzten Daten von Festplatten auf Magnetbänder, um teuren schnellen Plattenspeicher zu sparen, nach der Abkehr vom echten Magnetband für Online-Daten hielt sich noch die Produktbezeichnung „virtual tape library“. Und auch heute noch ist das von Herstellern so beworbene „All-Flash“-Prinzip nicht die wirtschaftlichste Variante, vorausgesetzt man beherrscht das Management verschiedener Speicherklassen.

Durch Backup/Restore, also der Sicherung und anlassbezogenen Wiederherstellung von Daten im Unternehmen, werden diese Datenmengen noch multipliziert: seriösen Schätzungen zufolge werden Daten aktuell etwa acht- bis zehnmal repliziert - hoffentlich ohne unbemerkte Änderungen - und auf den Speicher- und Backupsystemen bewegt und vorgehalten. Backup- und Archivsysteme sind nicht nur meist sehr umfangreich, sondern erfordern zusätzliche Infrastruktur für das Management und Komprimieren der Daten sowie volumen-gebundene Software-Lizenzen.

Klassifizierung – Umgang mit kritischen Daten, Metadaten, Klassifikation und Datenregeln

Je höher der Wert von Daten für ein Unternehmens, desto höher der Schutzbedarf - die kritischsten Daten erhalten den besten Schutz. Das erscheint logisch. Dazu muss das Unternehmen zunächst seine kritischen Daten kennen und wissen, was kritisch bedeutet: wie viel Arbeit samt erstellter Daten darf im Ernstfall verloren gehen, und wie lange darf es dauern, bis nach einem Ausfall die Applikation und die Daten wieder hergestellt sind? Fragen, die das Unternehmen nicht nur sich selbst und mit Hilfe seiner Dienstleister beantworten muss, auch bei Compliance Audits spielt das Thema eine große Rolle.

Vorsichtshalber alle Daten als kritisch zu behandeln, weil man es nun mal nicht genauer sagen kann, sollte hier nicht die Lösung sein. Die Spiegelung von Daten etwa zwischen ausreichend weit entfernten Rechenzentren in beinahe real-time ist sehr aufwändig und teuer, daher ist eine aktuelle und vollständige Klassifizierung der Daten wichtig.

Dateneignerschaft

In der Praxis scheitert dies immer wieder an der mangelnden Ausstattung der Fachbereiche mit benannten und aussagefähigen Dateneignern, modern Data Owner genannt.

Das Konzept der Dateneignerschaft und des Datenhaushalts klingt zwar zunächst einfach, die tatsächliche Ausgestaltung und Zuordnung von Verantwortung im Unternehmen erweist sich jedoch oft als herausfordernd. Da es aber ein wesentlicher Baustein einer erfolgreichen Data Governance ist, sollten wesentliche Element etabliert und gepflegt werden, wie

- Prinzipien und Rahmenwerk der Dateneignerschaft und des Datenhaushalts,
- Grundlagen der Entscheidungsfindung für Datenmanagementfragen und für die Erarbeitung einer Leitlinie Datenmanagement,
- Auflistung üblicher Herausforderungen beim Aufbau von Dateneignerschaft und Datenhaushalt und wie man ihnen begegnet,
- Bereitstellung bewährter Vorgehensweisen.

Der Data Owner ist per Definition die Person, die am Ende die Verantwortung für seinen Anteil an den Daten seiner Organisation und den daraus zu extrahierenden Informationen trägt. Er oder sie hat sich um Prüfung und Klassifizierung der Daten zu kümmern, um die Festlegung von Schutzbedarfen, Zugriffsregeln und -rechten, um Vorbereitung und Bearbeitung von Prüfungen, um angemessene Datensicherheit und -sicherung - all dies sind umfangreiche und wichtige Aufgaben.

Umgang mit Daten in historischen Formaten

In vielen Unternehmen gibt es noch archivierte Datenbestände in Formaten oder mit Zugriffmechanismen, die im Zuge der Migration auf neue Applikationen oder deren neuere Versionen nicht mitgenommen werden konnten. Gleichwohl müssen sie aus gesetzlichen oder vertraglichen Gründen noch aufbewahrt und im allerdings seltenen Notfall wieder gelesen oder gar aktiviert werden. Zu dem Zweck wird teilweise sogar noch die alte Hardware „im Keller“ aufgehoben, um Bandformate, CD-Stapel oder sogar Microfiche-Unterlagen wieder zum Leben zu erwecken.

Bedrohungen und Schutz von wichtigen Daten

Der Europäische Gerichtshof (EuGH) hat mit dem Urteil zum Fall Schrems II[1] klargestellt, dass europäische Daten in fremden Ländern – besonders der USA im Rahmen des Privacy Shield – nur gespeichert werden dürfen, wenn die Daten in diesen Drittländern einen mindestens gleichwertigen Schutz genießen, wie es die europäische Datenschutz-Grundverordnung (EU-DSGVO) innerhalb der Union vorsieht. Dies gilt umso mehr für Unternehmen der Finanz- und

[1] Vgl. https://www.security-insider.de/fragen-zur-eugh-entscheidung-im-fall-facebook-vs-schrems-a-949677/.

Versicherungsbranche sowie im Gesundheitswesen wegen sensibler Personen-bezogener Daten. Entsprechend hat der Europäische Datenschutzausschuss Handlungsempfehlungen[2] veröffentlicht, um mit der Rechtsprechung umzugehen und die neuen Anforderungen einzuhalten. Wer dies versäumt, dem drohen nicht nur empfindliche Strafen, sondern auch öffentliche Bloßstellung: Organisationen für die Durchsetzung des europäischen Datenschutzes, wie noyb, behalten Datenschutzsünder im Blick und stellten zuletzt 101 Unternehmen an den digitalen Pranger.[3]

Mangelnder Fokus auf IT-Sicherheit

Datenschutz beginnt nicht erst mit der sicheren Speicherung oder Verwendung von Daten, sondern bei der IT-Sicherheit. Oftmals fehlt es allerdings - vor allem in kleinen und mittelständischen Unternehmen - an Hilfsmitteln, wie z.B. Richtlinien und Werkzeuge für Kennwörter, definierten Zugriffsbeschränkungen und einer umfassenden Absicherung der Server. Wer allerdings den Weg zu den gespeicherten Daten nicht mindestens genauso absichert wie die Daten selbst, macht sich unter Umständen zur lukrativen Zielscheibe für Cyberkriminelle – und riskiert großen Schaden.

Datenschutz erfordert Zeit, Geld und kontinuierliche Aufmerksamkeit. Wegen der Entwicklung von Technologie und Systemen am Markt sowie der schnell fortschreitenden Digitalisierung von Unternehmen ist das Thema daher als dauerhafter Prozess anzusehen. Dies wird auch von den Aufsichtsbehörden durch immer schärfere Prüfungen der Konzepte und der praktischen Durchführung verdeutlicht.

Neue Daten-Herausforderungen in der Cloud oder business as usual?

Mittlerweile haben die meisten Unternehmen zumindest initiale Stufen einer Cloud-Migration durchlaufen, aber die zunehmende produktive Nutzung von Multi-Cloud und Data-Platform-as-a-Service (dPaaS) resultierte in letzter Zeit in noch häufigeren Migrationsaktivitäten, indem IT-Organisationen verstärkt auf der Suche nach den besten Cloud-Plattformen und Dienstangeboten sind.

Einer der wichtigsten Trends aus Sicht der Datenhaltung ist dabei, dass Cloud-native IT-Implementierungen die Art der Speicherplatzbereitstellung und des Zugriffs deutlich prägen und verändern werden. Auch die weiter oben schon angesprochene Data Governance, die schon in klassischen Umfeldern oft noch unvollständig ist, wird mit der Cloud nicht einfacher - im Gegenteil. Der Unterschied ist, dass der Cloudanbieter die Daten und üblichen Wege im Unternehmen nicht kennt und nicht aus eigenem Antrieb oder guter Gewohnheit die nötigen Arbeiten und Kontrollen durchführt, die oft niemand explizit spezifiziert und bestellt hat.

[2] Vgl. https://edpb.europa.eu/news/news/2020/european-data-protection-board-41st-plenary-session-edpb-adopts-recommendations_de.

[3] Vgl. https://noyb.eu/de/update-zu-noybs-101-beschwerden-ueber-eu-us-datentransfer.

Zugleich machen sich etwa 60% der Unternehmen ernste Sorgen um regulatorische Probleme, Governance und Compliance in Verbindung mit der Nutzung von Cloud-Diensten. Die Balance zwischen Aufbau und Unterhalt einer guten Daten-Governance zur Minderung oder Vermeidung von Risiken und gleichzeitig einer bestmöglichen Nutzung der Daten nehmen mehr als dreiviertel der Unternehmen zum Anlass, für die Cloudnutzung nach besser integrierten Datensicherheits- und Governance-Lösungen zu suchen als bisher.

Auch ist die Cloud nicht einfach ein anderer oder weiterer Ort, wohin man seine Datenverarbeitung verlagert, sondern es gibt vielfältige Geschäfts- und Betriebsmodelle, Multi-Clouds oder hybride Formen, wo Teile der Informationstechnik weiter im Haus sind, andere hingegen ausgelagert werden. Dies macht die Platzierung, Sicherung, das Management und das Synchron-Halten der Unternehmensdaten nicht einfacher.

Wo sind welche Instanzen der Daten – Placement und Discovery

Zum einen zeigt sich ein Trend, der Daten-Gravität genannt wird. An Orten, wo bereits durch Anbieter viele Daten gehalten werden, siedeln sich auch mehr Applikationen und Services an, und damit wieder weitere Daten. Durch diese technischen Ökosysteme werden sich solche Daten und ihre Verarbeitung an bestimmten Orten konzentrieren, wie in Deutschland z.B. im Raum Frankfurt, und sie sind von dort nicht mehr so leicht wirtschaftlich sinnvoll zu verschieben.

Andererseits wachsen mit den Datenmengen und ihrer Komplexität auch die Teams und die Anzahl der Datenquellen und Personen, die mit ihnen umgehen. Damit wird es immer schwieriger für den IT-Betrieb, Klarheit und Übersicht über all die Datentransformationen und -bewegungen zu behalten. Das beinhaltet auch die Speicherorte, von denen - siehe oben unter „Bedrohungen und Schutz von wichtigen Daten“ - nicht alle geeignet oder erlaubt sind, und die inhaltliche Validität der Daten je nach Quellen und Veränderungen.

Cloud Dienstleister oder Konkurrent?

Es gibt auch eine strategische Komponente: Firmen, die heute noch standardisierte IT-Services aus der Cloud anbieten, könnten sich morgen entscheiden, z.B. eine Fintech-Sparte aufzumachen und sind dann Servicegeber, Verwalter der Unternehmensdaten und Marktwettbewerber in einem. Ein Szenario, das bei den Hyperscalern in Teilen schon recht wahrscheinlich oder gar eingetreten ist.

Daten in historischen Formaten in der Cloud?

Im Inhouse-Betrieb konnte man mit archivierten Datenbeständen, die in historischen Formaten oder Zugriffmechanismen vorlagen, noch angemessen umgehen, solange die Wissensträger oder die Dokumentation ausreichte. In der Cloud wird es sehr schwierig, solche Daten abseits der gängigen modernen Datenstandards und -formate vorzuhalten. Es sei denn, es etablieren sich für diese Zwecke Nischenanbieter mit speziellen Angeboten zur Aufbewahrung oder Umformatierung von Daten, die dann aber sicher einen deutlichen Preisaufschlag gegenüber commodity services erfordern.

Datenwachstum – Datenspeicherung, -sicherung und -wiederherstellung

Wie schon ausgeführt sind diese Disziplinen in der „alten Welt“ schon herausfordernd genug. In der Cloud kann man sehr schnell aus einem ständig breiteren Angebot bestellen, man muss aber den Überblick und die Kontrolle behalten. So sind laut Anbieterstimmen Big Data und Speichervolumen im Petabyte-Bereich nur „einen Click“ entfernt, das wird aber in einem „pay per use“-Modell sehr schnell teuer - vor allem, wenn man Services und Volumina nutzt, die man nicht wirklich braucht.

Nicht zuletzt ist der Aspekt der klassischen Datensicherung möglicherweise existenziell für die Unternehmen. Bedrohungen für den Datenbestand können sich sowohl aus wirtschaftlichen Entwicklungen ergeben, auch wenn eine Problemlage heute unwahrscheinlicher ist, wie sie vor diversen Jahren um ein weit verzweigtes Geflecht von Cloudspeicherdiensten unter dem Dach eines Unternehmens mit dem bezeichnenden Namen „Nirvanix“ entstand. Die sprudelnden Quellen der Risikokapitalgeber versiegten, die Firma war sehr kurzfristig pleite und nicht mehr in der Lage, die großen Datenmengen wieder zu ihren Eigentümern zurückzuführen, was auch an fehlenden Bandbreiten und nicht mehr vorhandenen eigenen IT-Systemen bei ihren Kundenunternehmen lag.

Ein anderes Ereignis mit hohen Schäden ist noch sehr präsent: im März 2021 brannte ein moderner Rechenzentrumskomplex des europäischen Dienstleisters OVHcloud teilweise nieder, die benachbarten Rechenzentren mussten offline genommen werden.

Abbildung 1 - Brand des OVH-Rechenzentrums; © Feuerwehr Kehl

Optimisten könnten denken, bis auf die Downtime wäre das alles nicht so schlimm, die Daten sind doch in der Cloud immer noch irgendwo anders. In diesem Fall war das weit gefehlt, denn das sog. Shared-Responsibility-Modell heißt eben nicht, dass immer der Cloud-Provider verantwortlich ist, wenn Daten verloren gehen oder Sicherheitslücken auftreten. Der Cloud-Anbieter ist vielmehr zuständig für die Sicherheit seiner Cloud-Infrastruktur und der Basisprozesse, der Kunde hingegen für die Sicherheit seiner Applikationen und seiner Daten in der Cloud.

Das Angebot, die Daten auf alternativen Servern oder Public Clouds anderer Orte und Anbieter half im OVH-Fall nur den Kunden, die Ihre Daten anderweitig gesichert hatten. Wer sich nicht selbst um Datensicherungen gekümmert oder keinen Backup-Service gebucht hatte oder nur eine Vor-Ort-Sparversion in der gleichen Serverhalle, stand danach einem nicht mehr zu rettenden Datenverlust gegenüber und war ggfs. zusätzlich noch verpflichtet, eine hochpeinliche Meldung bei der jeweiligen Datenschutzbehörde abzugeben.

Im eigenen Unternehmen ist diese Stringenz zwar auch oft nicht durchgängig vorhanden, dort gibt es aber meist noch jemanden, der aus guter Gewohnheit und Mitverantwortung die notwendigen Schritte einleitet und dauerhaft durchführt. In der Cloud gibt es so jemanden nicht - in einem hoch standardisierten und automatisierten Ablauf muss man als Kunde selbst wissen, was man ordert und was nicht.

Drift von Metadaten, Klassifikationen und Datenregeln

So schlimm wie im Fall oben muss es nicht immer kommen. Eine eher schleichende Herausforderung für Datenowner und den IT-Betrieb ist es, die Strukturen und Qualität ihrer Daten in komplexen und dynamischen Serviceketten zu bewahren. Mit jedem weiteren Service oder der Parallelisierung auf Multi-Cloud-Instanzen wächst die Gefahr, dass die Regeln und Klassifikationen, Zugriffsrechte oder Metadaten für kritische Unternehmensdaten sich auseinanderentwickeln.

Ohne ein spezialisiertes Datenmanagement mit entsprechender Governance und fortschrittlichen Tools sind die rapide wachsenden Datenmengen, Datenquellen und Datenmodelle nicht mehr zu beherrschen. Außer der fehlenden Übersicht und Transparenz der Daten fehlt dann auch die Möglichkeit, Herkunft und Veränderung von Daten nachzuweisen und so die Compliance-Anforderungen zu erfüllen. Ohne Kontrolle über den Lebenszyklus der Daten - die sog. Data Lineage - kann man sich nicht mehr sicher sein, ob man diese Daten je nach ihrer Quelle verwenden darf und ob ihre Auswertung zu verlässlichen Informationen führt. Diese Risiken haben mit der Vielfalt und Flexibilität der Datenhaltung in der Cloud eindeutig zugenommen.

Fazit

Datenmanagement in der Cloud wird trotz aller neuen Entwicklungen nicht einfacher – die Verantwortung für wichtige Daten und Informationen verbleibt bei den Eignern der Geschäftsprozesse und der Daten im Fachbereich genauso wie im IT-Betrieb, ganz gleich wie tief gestaffelt und agil die Lieferkette gestaltet ist. Was in der Cloud nicht beauftragt wurde, tut auch niemand für Sie! Dies ist bei der erweiterten Nutzung von Cloud-Services erfolgskritisch und bei ganzheitlichem Ansatz zur Cloud-Migration sowohl in der Konzeption und Umsetzung als auch im laufenden Betrieb zu beachten. Unternehmen sind gut beraten, diese Aufgaben gründlich zu bearbeiten, damit die Antworten auf die Titelfrage beruhigend ausfallen können.

Kurt Tanneberger ist Senior Manager bei Eurogroup Consulting und Experte für IT-Architektur, IT-Service Management, Test- und Releasemanagement, IT-Infrastruktur und IT-Betrieb sowie IT-Programm- und Projektmanagement. Seine Beratungsprojekte beschäftigen sich u.a. mit der Einführung und Optimierung von Datenmanagement, mit Fragen wie Master Data, Storage und Backup, Archivierung und Housekeeping sowie Datenmanagement in der Cloud. Kurt Tanneberger ist Ingenieur und verfügt über 35 Jahre Erfahrung in der IT-Branche und Kundenberatung.

www.datanavigator.de ISSN 2702-1904

Über den DataNavigator

Unterschiedliche Autoren beleuchten in der Reihe DataNavigator die vielfältigen Facetten des Datenmanagements und zeigen Ihnen interessante Aspekte auf. Der DataNavigator wird von Eurogroup Consulting regelmäßig zweimonatlich herausgegeben und möchte Ihnen interessante Denkanstöße liefern. Wir freuen uns, Ihnen mit dem DataNavigator unsere Erfahrungen weiterzugeben und stehen Ihnen für die Diskussion gerne zur Verfügung! Sie erreichen Autoren und Herausgeber bei Anregungen und Feedback unter der E-Mail-Adresse datanavigator@eurogroupconsulting.de.

Bisher erschienen:

#1 Dr. Marco Weiß – *Welt und Wesen des Datenmanagements*; September 2020

#2 Mario Kubik – *„DLT" – Potenzial zur Zukunftstechnologie?*; November 2020

#3 Jonathan Misler – *Verwendung von Daten und Modellen in datengetriebenen Geschäftsmodellen*; Januar 2021

#4 Johannes Bänsch – *Data Analytics und die Cloud*; März 2021

#5 Kurt Tanneberger – *Ja, wo sind sie denn? – Daten im IT-Betrieb*, Mai 2021