

FOKUS: BI & ANALYTICS

Wachstum ohne Wildwuchs

Die IT muss den Wildwuchs bei Business Intelligence Tools verhindern, ohne das Innovationspotenzial von Big Data und Data Science zu schmälern. Das hier vorgestellte Modell zur Data Governance verspricht einen Lösungsansatz.

→ VON RÖGER MATHIS

Seit Neustem, also in Zeiten digitaler Transformation und Big Data, ist der Data Scientist der Star unter den Datenanalysten. Er wirft den dringend nötigen Blick auf die riesigen Datenmengen und kann diese auch interpretieren. Unter Zuhilfenahme entsprechender Tools integriert er Daten quasi on-the-fly und wird damit in vielen Fällen dem Bedürfnis der Fachabteilungen nach schnell verfügbaren Analysen gerecht.

So weit, so gut. Doch die stetig wachsende Leistung der Endgeräte und die grosse Zahl einfacher zu bedienender Tools hat in den Fachabteilungen die falsche Annahme aufkommen lassen, dass Datenanalysen heutzutage ein

Kinderspiel seien. Entsprechend fehlt es dem Business an Verständnis dafür, dass die IT-Abteilung für ihre Wünsche Zeit benötigt und die Umsetzung teuer werden kann.

SINNLOSES HIN UND HER

Die bestehenden Business-Intelligence(BI)- und Datawarehouse-Systeme (DWH) sind nur noch selten in der Lage, in nützlicher Frist und zu tiefen Kosten die wachsenden analytischen Anforderungen zu erfüllen, geschweige denn, diese in eine saubere und auf Dauer kostengünstige Infrastruktur zu integrieren. Immer häufiger ist deshalb zu beobachten, dass unter Umgehung der IT-Prozesse analytische Applikationen in Richtung Business verlagert werden.

Dabei waren es doch genau diese Inselösungen mit ihren unendlich komplexen Excel-Dokumenten und Access-Datenbanken, die

Unternehmen dazu gezwungen haben, mit gezielten BI-Initiativen und sauberen DWH-Infrastrukturen das Grundproblem in den Griff zu bekommen: Die Einzelentwicklungen und unkontrolliert wuchernden Tools in den Abteilungen konnte bald niemand mehr instand halten. Und nun haben wir wieder die gleiche Situation: Schwer zu wartende analytische Inselapplikationen entstehen planlos und können von niemandem mehr bedient werden, wenn der Datenguru der Abteilung gekündigt hat. Wenn sie nicht ganz nutzlos für das Unternehmen werden sollen, werden sie dann innerhalb eines Projekts von der IT integriert – mühsam und meist unter sehr hohem Aufwand.

Dieses Hin und Her gibt es, seit es IT-Systeme gibt: Die Fachabteilung kauft oder entwickelt Software, bis der Aufwand für den Unterhalt zu gross oder zu komplex wird, und

übergibt sie schliesslich der IT. Das führt nicht nur zu unnötigen Kosten, sondern auch zu Frust auf beiden Seiten. Andererseits lassen sich diese Vorgänge auch nicht einfach verbieten, sonst schwächt man die Innovationskraft im Unternehmen. Wie also das Problem in den Griff bekommen und gleichzeitig dem Business ermöglichen, schnell und agil aus Daten wichtige Erkenntnisse zu gewinnen?

DER NEUE LAYER: DATA LAB

Zunächst müssen auf Governance-Seite die verschiedenen analytischen Anwendungen eingefügt werden, um Ordnung in das Datenchaos zu bringen. Dazu hat IT-Logix ein logisches Data-Governance-Modell entwickelt. Es basiert auf einem Layer-Konzept mit folgenden Komponenten (vgl. Grafik):

■ **Quell-Layer:** Dieser Layer repräsentiert die Datenquellen selbst. Sie können interner wie externer Natur und sowohl strukturiert als auch unstrukturiert sein.

■ **Data Lake Layer (Eingangs-Layer):** eine Rohdatenkopie der Quelldaten.

■ **Governed Masterdata Layer:** Hier werden die wichtigsten Stammdaten des Unternehmens integriert, konsolidiert und qualitätsgesichert zur Verfügung gestellt.

■ **Governed Data Repository:** Dieser Layer integriert die wichtigsten Transaktionsdaten mit dem Governed Masterdata Layer und umfasst den Kernbereich qualitätsgesicherter und integrierter Daten.

■ **Governed Data Mart Layer:** Eine Ableitung aus dem Governed Data Repository, die eine neue, oft aggregierte Sicht auf die Daten repräsentiert.

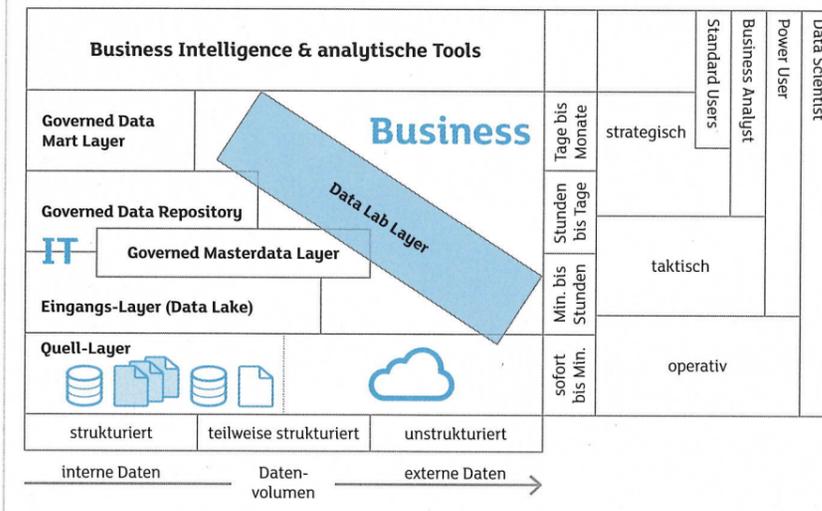
■ **Data Lab Layer:** Hier befinden sich die oben genannten Silolösungen, die nicht unter strikte IT-Governance fallen.

Dabei fällt der Data Lab Layer auf, den es im klassischen DWH-Modell nicht gibt. Er ist für die analytischen Inselapplikationen der Fachabteilungen da und erlaubt so Prototyping und Forschung im Bereich Data Science. Alle anderen Layer gehören ausschliesslich der IT und werden unter Berücksichtigung klar definierter Servicelevels professionell entwickelt und betrieben.

MERKMALE DEFINIEREN DEN ZWECK

Diese Services sind ein wichtiger Bestandteil des Modells. Ihnen werden pro Layer verschiedene Ausprägungen von Merkmalen zugewiesen. Das kann der Grad der Historisierung sein (z. B. Quellhistorie, mono/bi-temporal) oder die Performance, der Level der Datenintegration und -qualität, der Grad der Wiederverwendbarkeit oder die Datensicherheit. Als weitere Servicekriterien kommen etwa die Aktualität der Daten (in Echtzeit, minütlich, stündlich, täglich

Framework für die Data Governance



Mit dem Layer «Data Lab» erhält das Business Spielraum für Innovation

etc.), die Relevanz für Entscheidungsträger (strategisch, taktisch, operativ), die Qualität der Dokumentation oder die zu erwartete Projektdurchlaufzeit infrage.

So erhält der Eingangs-Layer einen viel tieferen Servicelevel als der Governed Data Mart Layer. Auf der höchsten Ebene wird nämlich davon ausgegangen, dass beispielsweise die Datenqualität und Performance hervorragend sind, die Daten jedoch nicht in Echtzeit zur Verfügung stehen. Services auf dieser Ebene werden eher strategischen Entscheidungsträgern Nutzen geben. Im Eingangs-Layer, wo weder Datenintegrationen noch Qualitätsprüfungen stattgefunden haben, die Daten jedoch in Echtzeit zur Verfügung stehen, werden hingegen die Daten eher operativen Entscheidungsträgern von Nutzen sein. Bilanzen, die eine 100-prozentige Datenqualität erfordern, werden beispiels-

Marketing zu finden, die beispielsweise bei Onlinemarketingkampagnen im Minutentakt Entscheidungen fällen und Aktionen einleiten müssen.

MEHR TRANSPARENZ, BESSERE STEUERUNG

Die konsequente Beschreibung der Services pro Layer legt versteckte Services und Aufwände offen und macht es so möglich, Servicelevel Agreements für analytische Applikationen abzuleiten. Durch die Einordnung der Anwendungen in das Layer-Konzept ist für das Management auch jederzeit transparent, welche analytischen Applikationen eigentlich im Unternehmen existieren, wo geforscht wird und wo Inseln bestehen, die zu einem späteren Zeitpunkt in die IT integriert werden müssen.

Des Weiteren beschreibt das Modell die Skills, die ein Data Engineer bzw. ein Datenkonsument auf den verschiedenen Layers benötigt. So wird z. B. ein Konsument, der auf den Eingangs-Layer zugreift, Skills im Bereich von SQL, Hadoop oder Python mitbringen müssen. Auf dem Governed Data Mart



«Das Modell lässt kontrollierte Freiheitsgrade für die Fachabteilungen zu»

Roger Mathis

weise auf dem Governed Data Mart Layer angesiedelt werden. Im Gegensatz dazu kann etwa eine Clickstream-Analyse direkt auf dem Eingangs-Layer aufgesetzt werden, da Datenqualität und -integration im Gegensatz zur Aktualität der Daten nicht im Vordergrund stehen. Allenfalls benötigt es hier eine Anreicherung mit Kundeninformationen, die aus dem Governed Masterdata Layer hinzugezogen werden können. Die Nutzer solcher Analysen sind dementsprechend auch eher im operativen

Layer hingegen ist es viel einfacher, Daten zu konsumieren. Unter Umständen muss der Nutzer hier nur ein einfaches Web-GUI oder eine App bedienen können.

Zudem lässt das Modell kontrollierte Freiheitsgrade für die Fachabteilungen zu, sodass Innovationen mittels Big Data und Data Science unterstützt werden können. Darüber hinaus lassen sich Fehlentwicklungen schnell korrigieren und unnötig hohe Kosten und Redundanzen vermeiden. ←

BILD: ISTOCKPHOTO.COM/SURKOVDMITRI

QUELLE: IT-LOGIX