

Testen im BI-Umfeld – ein Projekterfahrungsbericht

Mehr Effizienz mit Plausibilitätskennzahlen

Entscheidungen in Unternehmen basieren auf Informationen. Falsche Entscheidungen kosten viel Geld, häufig mehr als eine optimale Qualitätssicherung. Insbesondere bei Business-Intelligence-Lösungen sollte eine geringe Qualität in der Umsetzung nicht akzeptiert werden, da BI die Grundlage für eine Vielzahl von Entscheidungen bildet. Auswertungen müssen korrekte Ergebnisse erzeugen, sonst sind sie für die Wissensgenerierung und die anschließende Entscheidungsfindung unbrauchbar. Wir zeigen im Folgenden anhand eines realen Beispiels aus der Versicherungsbranche, dass eine effiziente Qualitätssicherung in BI-Projekten die Chance bietet, Projektkosten zu senken (siehe Abbildung 1) und die Risiken der Einführung eines BI-Systems deutlich zu reduzieren.

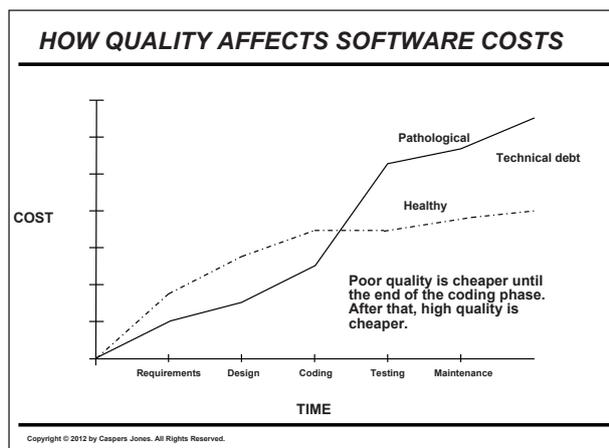


Abb. 1: Softwarequalität als Erfolgsfaktor – gezielte Qualitätsmaßnahmen senken die Gesamtkosten [Jon12]

Herausforderung: Testen im BI-Umfeld

Im transaktionalen Umfeld wird meist eine Benutzereingabe verarbeitet und nach entsprechenden Rechenoperationen ein Ergebnis zurückgegeben. Ein Beispiel wäre eine Point-of-Sale-Lösung für eine Autofinanzierung: Der Verkäufer selektiert ein Fahrzeug, legt Rabatte fest, sucht als Finanzierungsprodukt Leasing aus, trägt Parameter wie Laufzeit, Laufleistung und Anzahlung ein und stößt die Berechnung der monatlichen Rate an. Auch wenn die Rechenfunktion komplex sein mag – mit einem Taschenrechner und ein paar Excel-Zeilen lässt sich ein Soll-Wert für das Rechenergebnis definieren, der mit einem Test geprüft werden kann.

Im BI-Umfeld beschränken sich Benutzereingaben meist auf Selektionen von Darstellungsweisen und das Setzen von Filtern, die Verarbeitungsschritte sind jedoch viel komplexer. Es sind heterogene, unterschiedlich gut bekannte Datenquellen zu integrieren, die hinsichtlich der Datenqualität divergieren. Aus diesen Quellen werden Massendaten verschiedener Ausprägungen über mehrere Layer transportiert und schließlich aggregiert.

Hier stellt sich die Kernfrage: Wie ist überprüfbar, ob das angezeigte Ergebnis korrekt ist? Gegen welche Soll-Werte können die aggregierten Ergebnisse verglichen werden?

Die Bestimmung des Soll-Werts ist für eine Einzelsatzverarbeitung und überschaubare Datenmengen noch mög-

lich. Um zum Beispiel einen Soll-Wert für den Vertriebs-erfolg einer einzelnen Versicherungsagentur zu erhalten, können im Quellsystem die abgeschlossenen Verträge oder durchgeführten Beitragserhöhungen durchgezählt und wieder mit dem Taschenrechner oder ein paar Zeilen Excel ein Soll-Wert bestimmt werden.

Für Massendaten, die spätestens bei der Befüllung eines Data Warehouse verarbeitet werden müssen, scheitert dieser Ansatz. Ein manuelles Durchrechnen von Millionen von Datensätzen ist beinahe unmöglich. Woher weiß man also, ob bei rund fünf Millionen Verträgen im Bestand die im Management-Cockpit für 2012 ausgewiesenen Zahlen einer Gesamt-Beitragserhöhung von 7,3 Millionen Euro und einer Storno-Quote von 5,3 Prozent stimmen?

In der Praxis werden meist Vergleichszahlen aus bisherigen dispositiven Anwendungen herangezogen. Diese Methode hat jedoch Grenzen: Liegt wirklich die gleiche Datenbasis vor, oder wurden Altverträge nicht mehr berücksichtigt? Ist die Verarbeitungslogik noch immer die gleiche? Sind die Vergleichszahlen wirklich verlässlich oder möglicherweise selbst fehlerhaft? Und gibt es überhaupt ein System, das als verlässlich und korrekt gelten kann? Praxiserfahrungen zeigen: Experten stuften das aggregierte Ergebnis einer neuen BI-Lösung als falsch ein, weil es weder dem Wert X von Vergleichssystem A noch dem Wert Y von Vergleichssystem B entsprach – wobei X und Y stark voneinander abwichen und nicht feststand, ob nun X oder Y stimmte.

Soll-Werte für die BI-Anwendungen liegen nicht immer rechtzeitig zu Beginn des Komponententests vor. Deshalb hat sich für BI-Projekte zusätzlich eine detailliertere Referenzwertvorgabe bewährt, anhand derer schon während der ersten Komponententests die Datenversorgungsprozesse überprüft werden können.

Ein erster Schritt hierzu ist ein Referenzdatenbestand, der klar abgegrenzt und gut bekannt ist. Für diesen lassen sich klare Soll-Werte definieren. Was aber ist zu tun, wenn Realdaten verarbeitet werden sollen, die nicht im Detail bekannt sind und deren Menge eine Einzelbetrachtung unmöglich macht?

Eingangskontrolle

Eine wichtige, in vielen Bereichen selbstverständliche Praxis wird im BI-Umfeld oft übersehen: die gründliche Ein-

gangskontrolle. Mit der Eingangskontrolle ist kein Data Profiling gemeint, wie es im BI-Umfeld üblich ist, sondern eine Analyse der Quelldaten, um eine mess- und vergleichbare Aussage über die Plausibilität ihres Inhalts zu bekommen. Sinnvollerweise wird diese Kontrolle während der Projektlaufzeit bei jeder Datenanlieferung auf die Eingangsdaten durchgeführt, um eventuelle Fehler frühzeitig zu erkennen. Die Analyse erfolgt anhand im Vorfeld definierter Plausibilisierungskennzahlen, die nach verschiedenen Plausibilisierungsdimensionen ausgewertet werden.

Plausibilisierungskennzahlen sind einfache, aggregierbare numerische Attribute, die in den Quelldaten direkt abrufbar vorliegen. Was als Plausibilisierungskennzahl geeignet ist, hängt stark vom Anwendungsfall ab. Im Versicherungsumfeld sind zum Beispiel denkbar: Anzahl von Verträgen, Versicherungssummen, Einmal- und Regel-Beiträge.

Zwar lassen sich diese Plausibilisierungskennzahlen einfach über den gesamten Datenbestand aggregieren. Aussagekräftiger wird es jedoch, wenn sie nach verschiedenen Plausibilisierungsdimensionen ausgewertet werden, zum Beispiel nach Bewegungsarten wie Neugeschäft, Beitragsveränderung, Stormierung, nach Zeit oder – in Kombination mit weiteren Quellsystemen – nach Vertriebspartnern oder Durchführungswegen.

Die Plausibilisierungskennzahlen und -dimensionen sind im Idealfall mit dem Fachbereich oder einem Quellsystemspezialisten hinsichtlich ihrer Aussagekraft abgestimmt. Besonders vorteilhaft ist, wenn sie außerdem so definiert sind, dass sie mit den Soll-Werten in der BI-Anwendungsschicht verglichen werden können.

Die Ergebnisse der Eingangskontrolle einer Datenlieferung werden mit einem Quellsystemspezialisten auf Plausibilität geprüft. Ebenso kann schon zu diesem frühen Zeitpunkt festgestellt werden, ob die Soll-Werte des Fachbereichs grundsätzlich erreicht werden können.

Ist dies erfolgreich, gelten die Ergebnisse als Referenzwerte im weiteren Projektverlauf. Mit diesem Vorgehen können sowohl die weiteren Verarbeitungsschritte wie auch neue Datenlieferungen auf Plausibilität überprüft werden. Die Bearbeiter können zum Beispiel bei einer Ausweitung der Datenbasis auf weitere Zeiträume oder Kundengruppen schnell eine Einschätzung treffen, ob mögliche Abweichungen innerhalb tolerierbarer Größenordnungen anzusiedeln sind. Ist die Prüfung nicht erfolgreich, so beginnt die Fehlersuche an der richtigen Stelle – bei der Anlieferung, und dies, bevor ein einzelner Datensatz in das Data Warehouse geladen wurde.

Dies hat zwei Vorteile: Es wird keine überflüssige Zeit in das Laden der verschiedenen Layer der Datenhaltung in-

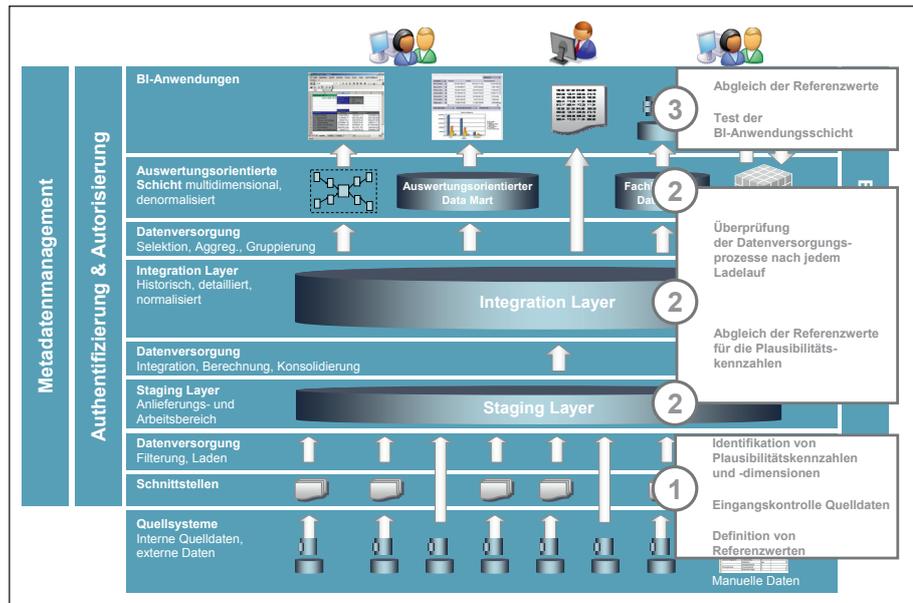


Abb. 2: Eingangskontrolle und Plausibilitätschecks – durchgängige Qualitätssicherung von der Anlieferung bis zur BI-Anwendung

vestiert, ebenso wird nicht mühevoll durch alle Layer nach der Stecknadel im Heuhaufen gesucht, die dort eventuell gar nicht zu finden ist. Abbildung 2 vermittelt einen Überblick über den Ablauf.

Die Ergänzung zu technischen und fachlichen Testfällen: Plausibilitätschecks

Nach erfolgreicher Eingangskontrolle und Referenzwertermittlung stoßen die Verantwortlichen die Befüllung der einzelnen Layer des Data Warehouse an. Nun gilt es, nach jedem Datenversorgungsprozess sogenannte Plausibilitätschecks durchzuführen. Damit wird geprüft, ob die Referenzwerte der Plausibilisierungskennzahlen, die bei der Analyse der Quelldaten ermittelt wurden, im nächsten Layer der Datenhaltung und schließlich in den BI-Anwendungen wieder erreicht werden. Auch dabei ist es hilfreich, mehrere Plausibilisierungsdimensionen zu überprüfen: So sollte zum Beispiel die Stornoquote bezogen auf die Angebote stichprobenhaft für eine einzelne Versicherungsagentur überprüft werden – möglichst bezogen auf Ebenen wie Produktgattungen oder Tarife.

Mit diesem Verfahren lassen sich Fehler schneller identifizieren, der bereits zitierte Heuhaufen, in dem die Stecknadel zu suchen ist, wird erheblich kleiner. Zudem liefern positive Plausibilitätschecks eine erste Aussage über positiv zu beurteilende Verarbeitungsschritte. Dies wirkt auch motivierend für das Projektteam: Wenn auch noch keine positiven Testfälle des Fachbereichs vorliegen, so sind zumindest die Datenversorgungsprozesse im Grundsatz korrekt.

Im vorliegenden Projekt haben die Beteiligten die Ergebnisse der Plausibilitätschecks im Verlauf des vorliegenden Projekts versioniert. So entstand eine Vergleichsbasis für weitere Tests, zum Beispiel bei erweitertem Quelldatenbestand oder für Regressionstests.

Plausibilitätschecks sind nicht geeignet, um die Korrektheit der gesamten Verarbeitungslogik zu überprüfen. Diese ist in der Regel zu komplex beziehungsweise über zu viele

Komponenten verteilt. Es empfiehlt sich, den Einsatz der Plausibilitätschecks projektindividuell festzulegen.

Werkzeugeinsatz und Automatisierung

Sowohl die Eingangskontrolle als auch die Plausibilitätschecks lassen sich mit Programmierungen, zum Beispiel SQL-Abfragen, oder individuell angefertigten Auswertungen durchführen. Sind die Referenzwerte und die Ergebnisse der Plausibilitätschecks in Vergleichstabellen abgelegt, sind Abweichungserkennungen einfach möglich.

Die Experten im Unternehmen können die genannten Aufgaben alternativ und meist komfortabler mit dem im BI-Projekt eingesetzten Technologie-Stack durchführen. Bei der Eingangskontrolle unterstützen dann Data-Profiling-Werkzeuge, mit passenden Plausibilitätsauswertungen lässt sich ein Berichtswesen bis hin zum Qualitätscockpit aufbauen.

Für Regressionstests ist es wichtig, die Eingangskontrolle sowie die Testläufe der Datenversorgungsprozesse automatisiert durchführen zu können. Wie praktisch, dass wir uns im BI-Umfeld bewegen: Automatisierte Datenversorgungsprozesse gehören zu den Hauptergebnistypen eines BI-Projekts. Für die Automatisierung der Eingangskontrolle, der Auswertungen und der Regressionstests wird zwar zusätzlicher Aufwand erforderlich, für den sich der Einsatz eines geeigneten Automatisierungs-Tools empfiehlt, die Automatisierung der Beladungsläufe selbst bietet aber erhebliche Synergieeffekte mit der eigentlichen Projektarbeit.

Einbindung in ein BI-Vorgehensmodell

Besonders vorteilhaft ist es, die dargestellten Aktivitäten in ein BI-Vorgehensmodell aufzunehmen. Dies ergänzt die bekannten Teststufen Komponententest, Integrationstests (inkl. nicht funktionaler, BI-spezifischer Tests wie zum Beispiel Speicherverbrauch, Performance), BI-Systemtest, fachlicher Systemtest und Abnahmetest (siehe Abbildung 3).

Basierend auf den gewonnenen Erfahrungen wurde im BI-Vorgehensmodell die „definition of done“ für den Komponententest wie folgt erweitert:

- Für die Eingangskontrolle der Quelldaten liegen Auswertungen vor. Ihre Ergebnisse sind zusammen mit den Quelldaten versioniert.
- Die Quelldaten sind geprüft. Ihre Plausibilität ist vom Quellsystemspezialisten bestätigt.
- Aus den Quelldaten sind Referenzwerte abgeleitet und versioniert.

Gerhard Bedners ist Mitarbeiter bei msg systems, München. E-Mail gerhard.bedners@msg-systems.com

	Komponententest	Komponentenintegrations-test	BI-System-Test	Fach-test	Last-tests	Abnahmetest
WAS						
Komponente	■					
Integrierte Komponenten		■	■	■	■	■
WER						
Entwickler	■					
Testteam msg	■	■	■	■	■	■
Testteam Kunde				■	■	■
WO						
Entwicklungsumgebung	■					
Funktionstestumgebung		■	■	■	■	■
Gesamt-Integrationstestumgebung		■	■	■	■	■
Lasttestumgebung (spez. Clients)					■	

*Gesamt-Integrationstestumgebung: Enthält alle lauffähigen Umsysteme.

Abb. 3: Analytisches Testen – vom Test der Einzelkomponente über das integrierte Gesamtsystem zum Abnahmetest

- Für die Komponente liegen versionierte Testfälle und Testdaten vor.
- Für die Komponente liegen versionierte Plausibilitätsauswertungen vor.
- Die Testfälle und die Plausibilitätschecks sind erfolgreich durchgeführt. Die Ergebnisse sind versioniert.
- Für die Komponenten und deren Plausibilisierungstests wurde eine Test-Automatisierung erstellt.

Fazit

Insbesondere im BI-Umfeld sind hohe Qualität und eine effiziente Testdurchführung gefragt. Hierfür sollten die Vorgehensmodelle entsprechend spezialisierte Testaktivitäten beinhalten. Start des Vorgangs ist die Eingangskontrolle der Quelldaten, gefolgt von einer Referenzwertbestimmung gemeinsam mit dem Fachbereich als Arbeits- und Testfundament. Darauf basierende Plausibilitätschecks und -auswertungen sichern die Korrektheit der Datenversorgungsprozesse ab oder decken Abweichungen auf. Automatisierte Datenversorgungsprozesse schaffen Wiederholbarkeit insbesondere für Regressionstests. Die dadurch mögliche Konzentration auf die Fachlichkeit ist ein effektiver Beitrag zur Reduzierung von Aufwand und Risiko und sichert letztlich den angestrebten Projekterfolg.

[Literatur]

[Jon12] Jones, C.: Software Quality in 2012: A Survey of the state of the art. <http://sqqne.org/presentations/2012-13/Jones-Sep-2012.pdf>, abgerufen am 22.7.2013

BI-SPEKTRUM ist eine Fachpublikation des Verlags:
SIGS DATACOM GmbH | Lindlaustraße 2c | 53842 Troisdorf
Tel.: +49 (0) 22 41.2341-100 | Fax: +49 (0) 22 41.2341-199
E-mail: info@sigs-datacom.de
www.bi-spektrum.de

SIGS DATACOM
FACHINFORMATIONEN FÜR IT-PROFESSIONALS

Kurzprofil msg

msg ist eine unabhängige, international agierende Unternehmensgruppe mit weltweit mehr als 4.500 Mitarbeitern. Mit ihrem ganzheitlichen Leistungsspektrum aus strategischer Beratung und intelligenten IT-Lösungen hat sie sich in über 30 Jahren einen ausgezeichneten Ruf als Branchenspezialist erworben.

Services und Solutions

msg bietet umfassende Beratung und Unterstützung bei allen Fragen rund um Business Intelligence. Definierte strategische und technische BI-Services sowie fachliche BI-Solutions erlauben eine professionelle und zielgerichtete Unterstützung unserer Kunden.

Das BI-spezifische Vorgehensmodell msg.PROFI-BI erlaubt die systematische Erbringung unserer BI-Services und -Solutions. Branchenspezifische Frameworks für fachliche Fragestellungen – beispielsweise zum Kunden-, Vertriebs-, Risiko-, Asset-, Geschäftsprozess- und Performancemanagement – liefern die **fachlichen BI-Solutions**. Die **technischen BI-Services** adressieren Anforderungen zu bereits bestehenden oder neu zu etablierenden BI-Landschaften. Schwerpunkte unserer **strategischen BI-Services** sind beispielsweise die BI-Strategie-, BI-Organisations-, BI-Architekturberatung und Themen wie BI-Software-Auswahl und Steuerung der Datenqualität.

Von der BI-Strategieberatung über die Konzeption und Umsetzung der optimalen BI-Architektur sowie der Auswahl passender BI-Plattformen und -Werkzeuge – beispielsweise zum Data Mining –, dem Einsatz geeigneter BI-Solutions bis hin zur Unterstützung im laufenden Betrieb bietet msg qualifizierte Leistungen aus einer Hand.

Referenzen

Die richtige Information zur richtigen Zeit am richtigen Ort ist die Basis für operative, taktische und strategische Entscheidungen. Unsere Services und Solutions machen dies möglich. Das wissen auch unsere Kunden, zu denen Unternehmen aus verschiedenen Branchen, wie Automotive, Finanzdienstleistungen, Versicherungen, Industrie sowie Life Science & Healthcare zählen.

Ausgewählte Referenzen:

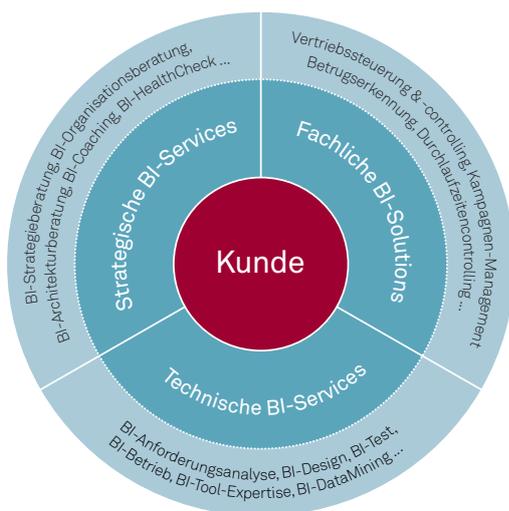
Allianz, Daiichi Sankyo, Munich Re, Siemens AG, VW Financial Services

Partner

Verlässlichkeit, Wertschätzung und Vertrauen – darauf basiert die Zusammenarbeit mit unseren Mitarbeitern, unseren Kunden und unseren Partnern. Wir arbeiten mit ausgewählten Unternehmen, teilweise schon seit vielen Jahren, eng zusammen.

Wichtige Partner:

IBM, Information Builders, Microsoft, SAP, SAS



Auszug aus unserem Business-Intelligence-Portfolio



msg systems ag

Robert-Bürkle-Straße 1 | 85737 Ismaning/München

Telefon: +49 89 96101-0 | Fax: +49 89 96101-1113

www.msg-systems.com | info@msg-systems.com

