

Datenqualität erfolgreich steuern

Neu in RapidRep 5.0:
Datenvergleichsassistent

Die RapidRep Test Suite ist eine innovative Software für das automatisierte Testen des Back-Ends.

RapidRep kann auf Datenbestände unterschiedlicher Systeme zugreifen und diese beliebig miteinander verknüpfen. Zu den unterstützten Datenquellen zählen alle gängigen Datenbanken (Oracle, DB2, Microsoft Access, Sybase, SQL Server u.v.m.) sowie strukturierte Textdateien.

Die Flexibilität von RapidRep zeigt sich in den zahlreichen Einsatzmöglichkeiten. RapidRep kann z.B.:

- die Ergebnisse einer Datenverarbeitung verifizieren.
- die korrekte Durchführung einer Migration nachweisen.
- die Qualität von Daten (Rohdaten und Ergebnisse) messen und verbessern.
- Tests automatisiert wiederholen und die Ergebnisse an ein Testmanagementsystem übertragen.

RapidRep wird von FINARIS entwickelt, vertrieben und kontinuierlich verbessert.

Entsprechend Ihrer Einkaufsstrategie können Sie die RapidRep Test Suite kaufen oder mieten.

Unser aktuelles Schulungsangebot finden Sie auf unserer Website. Bei Bedarf führen wir speziell auf Ihre Bedürfnisse zugeschnittene Workshops bei Ihnen oder uns durch.

Testmanagement

Unternehmen nutzen Testmanagementsysteme (TMS) zum Planen, Verwalten und Nachverfolgen von Testfällen. Das Testen des Back-Ends wird von diesen Systemen nur sehr eingeschränkt unterstützt. RapidRep hingegen wertet Testfälle maschinell aus und dokumentiert das Ergebnis in einem der unterstützten Systeme wie HP Quality Center, IBM Rational Quality Manager oder Microsoft Team Foundation Server. Für jeden durchgeführten Testfall erstellt RapidRep eine detaillierte Excel Arbeitsmappe. Im TMS erbringen diese Arbeitsmappen den reversionssicheren Testnachweis.

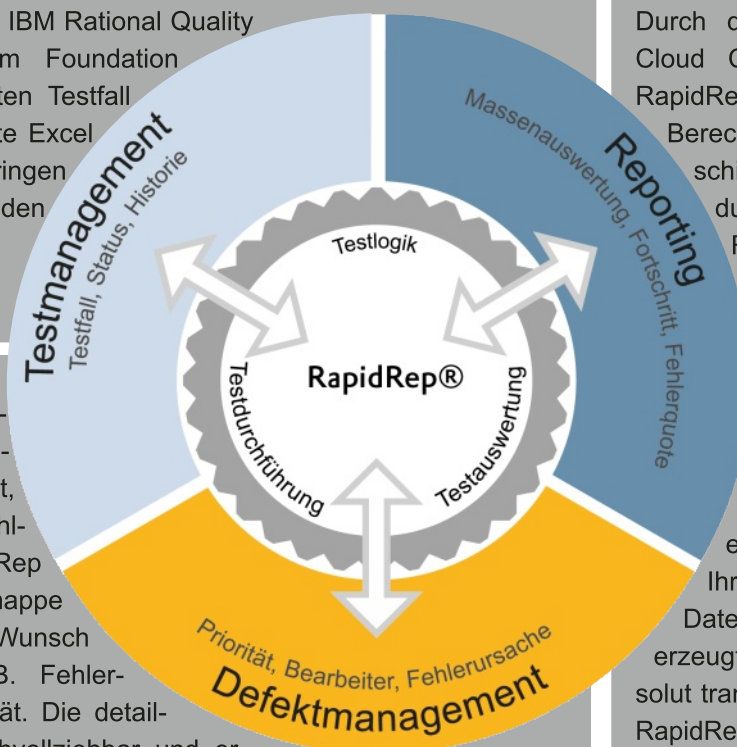
Fehlermanagement

Stellt RapidRep bei der Testauswertung mindestens eine Abweichung vom Soll-Ergebnis fest, legt es einen Defekt zum fehlgeschlagenen Testfall an. RapidRep fügt die erzeugte Excel Arbeitsmappe hinzu und ergänzt je nach Wunsch weitere Informationen wie z.B. Fehlerursache, Bearbeiter oder Priorität. Die detaillierten Arbeitsmappen sind nachvollziehbar und erleichtern die Fehlersuche. RapidRep unterstützt zahlreiche Fehlermanagementsysteme.

Reporting

RapidRep kann Ergebnisse und Zwischenergebnisse der Testdurchführung in einer Datenbank abspeichern und anschließend auswerten. Durch den möglichen Einsatz von Cloud Computing Technologien ist RapidRep in der Lage, sehr viele Berechnungen parallel auf unterschiedlichen Datenbeständen durchzuführen.

RapidRep stellt die Analyse-Ergebnisse in benutzerdefinierten Excel Arbeitsmappen dar und verschafft Ihnen so den Überblick über sehr große Datenbestände. Beispiele für solche Massenauswertungen sind etwa Berichte über den Stand Ihrer Datenmigration oder Datenqualitätsauswertungen. Die erzeugten Arbeitsmappen sind absolut transparent und nachvollziehbar. RapidRep legt die Reports reversionssicher in einem Repository ab.



Betriebliche Prozesse benötigen Geschäftsdaten von hoher Qualität. Aktuell, konsistent und widerspruchsfrei sollen sie sein. Doch die Realität sieht oft anders aus. Zunehmend verstehen Unternehmen Datenqualität daher als Prozess und als strategisch wichtige Aufgabe, in die es sich zu investieren lohnt.

Mit Hilfe von RapidRep können Unternehmen branchenunabhängig die Qualität ihrer Geschäftsdaten nachhaltig und kostengünstig verbessern.

Daten müssen sowohl technische als auch inhaltliche Anforderungen erfüllen. Daher ist es von Vorteil, wenn bei der Verbesserung der Datenqualität technisch und fachlich ausgerichtete Mitarbeiter eng zusammenarbeiten.

Ein wirksamer Prozess zur Verbesserung der Datenqualität muss daher in der Lage sein, Mitarbeiter aus Fach- und IT-Abteilung einzubinden. Ein solcher Prozess setzt sich grundsätzlich aus vier Phasen zusammen (siehe nachfolgende Grafik).



Definition der Datenanforderungen

Mitarbeiter aus dem Datenqualitätsteam legen fest, welche Datenquellen überprüft und welche Eigenschaften die Geschäftsdaten haben müssen (sog. Invarianten). Es kann aber auch hilfreich sein, genau jene Konstellationen zu beschreiben, die fehlerhafte oder unvollständige Daten charakterisieren.

Als allseits akzeptiertes und geeignetes Werkzeug bietet sich Excel für die Festlegung der Regeln und zur Kommunikation zwischen Mit-

arbeitern aus IT- und Fachabteilung an.

In Excel definieren Mitarbeiter, ...

... welche Daten (Tabellen, Abfragen oder Dateien) getestet werden sollen.

... welche Prüfungen (Tests) auf den Daten angewandt werden und wie die dazugehörigen Fehlerhinweistexte lauten, falls Datensätze die Prüfungen nicht bestehen.

... Schwellenwerte für die Anzahl

Regel #	Prüfung	Vorbedingung	Zusatzbedingung	Folgerung
1815	PLZ1	LAND = 'DE'	LÄNGE(PLZ) <> 5	FAILED
1816	PLZ1	LAND = 'DE'	PLZ IN (#DOMÄNE('DE-PLZ'))	PASSED
1817	PLZ1	LAND = 'DE'	-	FAILED
1818	PLZ1	Rest	PLZ IS NULL	FAILED
1911	T1_ISIN	LÄNGE(ISIN) <> 12	-	FAILED
1912	T1_ISIN	Rest	#ISINChecksum(ISIN) <> OK	FAILED

fehlerhafter Datensätze, ab deren Überschreiten eine Benachrichtigung erfolgen soll.

... Regeln für die einzelnen Prüfungen, die RapidRep dazu verwendet, um fehlerhafte Datensätze ein-

deutig identifizieren zu können.

Am Beispiel zur Überprüfung einer Postleitzahl wird die Vorgehensweise anhand eines kleinen Regelwerkes erläutert (siehe Tabelle oben):

1. RapidRep ermittelt alle Regeln, die für die Prüfung, in diesem Fall 'PLZ1', existieren. In unserem Beispiel sind das die Regeln 1815 bis 1818. RapidRep wertet Regeln immer in aufsteigender Reihenfolge aus.

2. Zu jeder Regel überprüft RapidRep, ob der untersuchte Datensatz sowohl die Bedingung in der Spalte „Vorbedingung“ als auch in „Zusatzbedingung“ erfüllt. Falls ja, dann endet hier die Auswertung für diesen Datensatz. Das Ergebnis leitet sich von der Spalte „Folgerung“ (PASSED oder FAILED) ab.

Die Unterscheidung in „Vorbedingung“ und „Zusatzbedingung“ erhöht die Lesbarkeit. Fallunterscheidungen und Restmengen lassen sich auf diese Weise gut erkennen.

3. Falls Vorbedingung oder Zusatzbedingung nicht erfüllt sind, wertet RapidRep die nächste Regel analog zu Punkt 2 aus. Falls keine zutreffende Regel existiert, gilt der Datensatz als valide.

Die untenstehende Tabelle stellt das Ergebnis der PLZ-Prüfung am Beispiel von fünf Datenkonstellationen dar. RapidRep stellt bei drei Datensätzen einen Regelverstoß fest.

LAND	PLZ	Regel	Resultat
'DE'	'99999'	1817	FAILED
'DE'	'60594'	1816	PASSED
'DE'	'1129'	1815	FAILED
'CH'	leer	1818	FAILED
'GB'	'SE1 7TJ'	-	PASSED

Die vorgestellte Modellierung besticht durch ihre Klarheit und Transparenz.

Die Regeln sind verständlich, eindeutig und flexibel erweiterbar. Beim Aufbau von Regeln können Makros verwendet werden, wie in diesem Beispiel `#DOMÄNE('DE-PLZ')` zur Abfrage von gültigen Werten für deutsche Postleitzahlen.

Mit einem ähnlichen Ansatz lassen sich auch Aussagen über Mengen von Datensätzen treffen. Entsprechende Regeln können z.B. feststellen, ob genau ein Datensatz pro Jahr und Kunde existiert, ob ein bestimmtes Datenbankfeld sowohl positive als auch negative Werte enthält oder ob Wertebereiche in Feldern wie erwartet variieren.

Messungen durchführen

Die definierten Regeln bilden die Logik zur Messung der Datenqualität. Auswertungen über die Verteilung von Datenausprägungen im Zeitablauf können zur Plausibilisierung verwendet werden. RapidRep führt Messungen maschinell und vollständig automatisiert durch. Zur kontinuierlichen Qualitätsmessung lässt sich RapidRep periodisch (z.B. täglich), zeitgesteuert oder individuell parametrisiert ausführen. RapidRep speichert alle fehlerhaften Datensätze in einer Datenbank.

Parameter beim Aufruf von RapidRep steuern den Umfang der Auswertungen und kennzeichnen die Ergebnisse einer Messung, um Veränderungen gegenüber vorherigen Auswertungen genau verfolgen zu können.

Ergebnisse analysieren

Nun gilt es aus den vielen Messergebnissen Erkenntnisse abzuleiten, die später als Grundlage für Verbesserungsmaßnahmen dienen können.

RapidRep stellt zu diesem Zweck die Ergebnisse von Messungen in einer detaillierten Excel Arbeitsmappe dar.

Das Übersichtsblatt listet die Anzahl fehlerhafter Datensätze pro Prüfung auf und stellt sie den erfolgreich geprüften Datensätzen gegenüber.

Das Delta-Arbeitsblatt stellt die Veränderungen gegenüber vorherigen Analyseläufen dar. Übersteigen diese einen absoluten oder relativen Schwellenwert, erfolgt eine farbliche Kennzeichnung (siehe Abb.).

Datenquelle	Prüfung	# Fehler (aktuell)	# Fehler (vorher)	Fehlertext
ADRESSEN	PLZ1	148	150	PLZ fehlt oder ungültig
CLEARING	T1_ISIN	9	5	ISIN nicht korrekt

Zu jeder geprüften Datenquelle existiert ein separates Arbeitsblatt, das die fehlerhaften Datensätze zu dieser Datenquelle auflistet.

Ein weiteres Arbeitsblatt listet die aktuell gültigen Datenqualitätsregeln auf, die RapidRep zur Auswertung herangezogen hat. Die Fehlerbewertung ist durch die

Korrektur-Regel #	Verstoß gg. Regel #	Vorbedingung	Zusatzbedingung	Korrektur
7099	1815	LÄNGE(PLZ) = 4	'0' & PLZ IN (#DOMÄNE('DE-PLZ'))	UPDATE ADRESSEN SET PLZ = '0' & PLZ

kompakte und verständliche Darstellung auch Mitarbeitern ohne technisches Spezialwissen möglich.

Es kommt in der Praxis häufig vor, dass auffällig gewordene Datenkonstellationen von der nächsten Messreihe ausgeschlossen werden sollen. Zu diesem Zweck existiert eine Ausschlusstabelle, deren Einträge beispielsweise verhindern, dass RapidRep altbekannte und

akzeptierte Datenschiefe immer wieder ausweist.

Mängel beseitigen

Zum Aufspüren von Fehlern hilft nur intensives Testen. Die Ergebnisse, die ein Programm erzeugt und abspeichert, können mit dem hier vorgestellten Ansatz über Regeln plausibilisiert werden.

Um bereits fehlerhafte Daten zu verbessern, besteht die Möglichkeit zur Definition von Korrekturregeln. Dabei wird die Tatsache ausgenutzt, dass RapidRep die eindeutige Regel-ID kennt, gegen die ein

Datensatz verstoßen hat. Für jede Regel-ID kann das Datenqualitätsteam festlegen, welche Gegenmaßnahmen in Form von Updates, Default-Werten etc. ergriffen werden können.

Beispiel: Ein Datensatz mit dem Wert '1129' (Dresden) im Feld PLZ wird über ein fehlerhaftes Liefer-

system ohne führende Null (0) angeliefert und verstößt daher gegen die Regel 1815.

Der Korrekturmechanismus überprüft mit Regel 7099, ob die Postleitzahl 4-stellig ist und ob '01129' eine gültige Postleitzahl wäre. Falls ja, dann wird der Datensatz um die fehlende Null ergänzt.

Die hier am einfachen Beispiel vorgestellte Lösung ist für beliebig komplexe Datenprüfungen geeignet.

RapidRep konnte in verschiedenen Projekten die Kosten beim Testen des Back-Ends deutlich reduzieren. Die mit der Kostenersparnis einhergehende Qualitätssteigerung ist eine direkte Konsequenz der folgenden Eigenschaften von RapidRep.

Automatisierung

RapidRep verfügt über ein leistungsfähiges Batch-Interface, das die Auswertung von Testfällen und die Erzeugung von Reports hochgradig automatisierbar macht. Dabei erledigt RapidRep die zugrundeliegenden Berechnungen ohne jegliche manuelle Zwischenschritte. Die Auswertung des gesamten Testfallportfolios lässt sich beliebig oft wiederholen.

Hohe Integration

RapidRep kommuniziert mit allen wichtigen Test-/Fehlermanagementsystemen. So lässt sich RapidRep reibungslos und ohne manuelle Zwischenschritte transparent in einen übergeordneten Testprozess einbinden. Für den RapidRep-Anwender sind keine Kenntnisse über das dahinterliegende Test- bzw. Fehlermanagementsystem notwendig.

Effizientes Einbinden von Ressourcen

Bei RapidRep spielen Excel Arbeitsmappen eine wichtige Rolle. Diese ermöglichen es, Mitarbeiter aus Fach- und IT-Abteilung gemeinsam und frühzeitig in die Gestaltung von Excel Vorlagen einzubinden. Dadurch ist sichergestellt, dass die im Verlauf der Testauswertung automatisch befüllten Arbeitsmappen sämtliche zum Nachvollziehen der durchgeführten Tests erforderlichen Details enthalten. Der Hauptaufwand liegt damit nur noch in der Fehleranalyse und nicht mehr in der ansonsten aufwendigen Durchführung und Dokumentation der Tests.

Global Delivery Model

RapidRep gewährleistet dank seiner modularen Architektur aus unterschiedlichen, spezialisierten Komponenten die organisatorische Trennung von Testfalldesign und Testdurchführung. Außerdem legt RapidRep sämtliche Artefakte versioniert in einem Repository ab. So können unterschiedliche Personen von überall nachvollziehbare und konsistente Änderungen an der Testimplementierung vornehmen.

Wiederverwendbarkeit

Viele Bausteine einer Testimplementierung lassen sich in anderen Projekten wiederverwenden. So wachsen die mitgelieferten Funktionsbibliotheken zu einer Sammlung für immer wiederkehrende Aufgaben heran. Für folgende Standardprobleme stehen Ihnen bereits fertige Lösungen zur Verfügung:

- Wizard zum Abgleich von Daten aus unterschiedlichen Quellsystemen (Datenvergleichsassistent)
- Wizard für regelbasierte Datenqualitätsauswertungen

Beide Komplettlösungen sind nach einmaliger Konfiguration sofort einsatzfähig.

Alleinstellungsmerkmal

Auf dem Markt für Testautomatisierung gibt es derzeit keine mit RapidRep vergleichbare Lösung. Die nachweisbaren Projekterfolge zeigen, dass sich das Testen des Back-Ends mit RapidRep hochgradig automatisieren lässt. Die Lizenzkosten zum Kauf oder zur Miete von RapidRep amortisieren sich in kurzer Zeit.

HP Quality Center, IBM Rational Quality Manager, Microsoft Excel, Microsoft Team Foundation Server, Oracle, DB2, Sybase, SQL Server und RapidRep sind (eingetragene) Marken der jeweiligen Inhaber.

Systemvoraussetzungen



- 1 GB RAM, 400 MB Plattenspeicher
- Windows XP SP3, Vista, 7 und 8
- RedHat- und SUSE-Linux GTK > 2.2.1
- IBM AIX
- Solaris / Open Solaris

Evaluierung / Proof of Concept



Möchten Sie RapidRep® an einem spezifischen Anwendungsfall in Ihrem Unternehmen evaluieren?

Sprechen Sie uns an.

Wir realisieren gerne eine RapidRep Testinstallation und überzeugen Sie persönlich von den Vorteilen unserer Software.