

Datenintegration unter verschärften Bedingungen:

# Migration und Synchronisation für Cloud-Applikationen

Dr. Daniel Rolli, Conemis AG

*Wie bedient man effizient die Integrationsschnittstellen von Cloud-Applikationen, wenn der Betreiber die Schnittstellen, Datenformatierung, Speicherplatz oder das tägliche Übertragungsvolumen für Daten reglementiert? Web Service Allotments sind hier nur ein gegenwärtiges Beispiel, das diese Fragestellung bei einigen Kunden von Oracle verschärft hat. Dieser Artikel zeigt die Herausforderungen bei der Datenintegration für Cloud-Applikationen am Beispiel von Oracle CRM On Demand.*

Seit Jahrzehnten werden Daten zwischen Systemen abgeglichen und aus alten Software-Applikationen an neue übergeben. Jede IT-Abteilung kennt das. Cloud Computing mit Software as a Service (SaaS) bringt dabei in Teilen grundlegend neue Herausforderungen mit sich. In Zeiten, in denen Cloud und SaaS auf die Realität der IT-Operations treffen und die etablierten Software-Anbieter mit ihren eigenen Cloud Applications zu den Herausforderern in den Ring gestiegen sind, beschleunigt sich die Weiterentwicklung von Schnittstellen. Oracle positioniert hier prominent seine Fusion Applications, hat aber mit Applikationen wie Oracle CRM On Demand (OCOD) schon heute den Einstieg in die Cloud- und SaaS-Welt realisiert. Technisch propagieren die Anbieter Standards und Quasi-Standards wie „Soap“, „WSDL“, „REST“ etc. und kultivieren doch ihre eigenen Optimierungen sowie Anreicherungen dieser Standards. Konzeptionell legen sie bei Architektur und Zuschnitt von Schnittstellen ganz unterschiedliche Paradigmen zugrunde und wirtschaftlich treibt eine beginnende Rentabilitätsorientierung die Reglementierung von Kapazitäten bei Speicherplatz, Rechenleistung und Bandbreite für Cloud-Applikationen voran.

## Szenarien für die Datenintegration

Die Szenarien „Einführung“, „Konsolidierung“, „Synchronisation“ und „Anbindung von Applikationen“ in Abbildung 1 veranschaulichen die Anwendungsfälle für Datenintegration sowie das Verständnis von Datenintegration, auf dem dieser Artikel

basiert. Die hier beschriebenen Szenarien sind zunächst unabhängig davon, ob Cloud- oder herkömmliche Applikationen betrachtet werden.

Bei der Einführung neuer Applikationen unterstützt die Datenmigration aus Altsystemen den Wechsel zur neuen Applikation. Sie ergänzt die Beratung des Kunden sowie die Konfiguration der neuen Applikation und ist eine zeitlich begrenzte Datenintegrationsleistung, die pro Ziel-Instanz der Applikation meist nur einmal erbracht wird. Sie ermöglicht es dem Kunden, in der neuen Applikation unter Zugriff auf alle bestehenden Daten nahtlos weiterzuarbeiten. Im Fall von Oracle CRM On Demand (OCOD)

hat der Kunde nach der Einführung seine bestehenden Accounts, Contacts, Dokumente etc. direkt verfügbar.

Die Konsolidierung von Systemen ist vor allem bei Unternehmens-Übernahmen und -Fusionen notwendig, aber auch bei der Optimierung der IT-Landschaft innerhalb eines Unternehmens. In der Regel ist die Konsolidierung wie die Migration in eine neu eingeführte Applikation eine einmalige Leistung, die den Umzug von Daten beinhaltet. Die Datenkonsolidierung wird deshalb als Form der Migration mit zusätzlichen Anforderungen betrachtet. Im Kontrast zur einfachen Migration bringt sie die Notwendigkeit mit sich, dass Daten aus den



Abbildung 1: Szenarien Datenintegration

Quellsystemen mit denen im Zielsystem zusammgeführt werden. Das muss zudem oft bei parallelem Produktivbetrieb von Quell- und Zielsystemen erfolgen.

„Datensynchronisation“ bezeichnet den regelmäßigen Abgleich von Daten zwischen Applikationen. Dieser erfolgt entweder von einem Master-System hin zu einer oder mehreren Applikationen oder kann bidirektional zwischen Applikationen stattfinden. Bei einer Synchronisation mit hoher Frequenz, zum Beispiel mehrfach pro Tag, ist es unerlässlich, dass der Datenabgleich nach Einrichtung vollautomatisiert erfolgt.

„Anbindung“ bezeichnet den Datenaustausch zwischen unternehmensinternen und externen Systemen. Wenn hierbei, wie in den meisten Fällen, die Übertragung regelmäßig erfolgt, wird die Anbindung als Form der Synchronisation mit besonderen Anforderungen betrachtet. Für den organisationsübergreifenden Datenaustausch sind dabei explizit festgelegte und dokumentierte Austauschformate notwendig. Sie ermöglichen auch bei unterschiedlichem internen Aufbau von Software und Datenmodellen die verständliche Kommunikation zwischen Applikationen.

Bemerkenswert ist, dass die Einbindung von Cloud-Applikationen die Grenzen zwischen einer unternehmensinternen Synchronisation und einer externen Anbindung verwischt. Der Kunde sollte bei einer für ihn bereitgestellten Cloud-Applikation zwar die alleinige Hoheit über die darin enthaltenen Daten haben, es bedarf jedoch explizit definierter Austauschformate wie bei einer externen Anbindung. Diese können nicht vom Kunden allein bestimmt und meist auch nur in engem Rahmen angepasst werden. Sie benötigen eine übergreifende Einigung zwischen Kunde und Anbieter.

### Herausforderungen in der Cloud

Abbildung 1 führt neben den Szenarien die drei zentralen Herausforderungen für jede Datenintegration auf: „störungsfreier Geschäftsbetrieb“, „konsistenter und vollständiger Datenbestand“ und „Qualitätssicherung“. Diese gelten unabhängig davon, ob die Integrationsanforderungen Cloud-Applikationen einschließen oder nicht. In Cloud-Szenarien ist der Wandel von Schnittstellen, getrieben durch die eingangs genannten Einflussfaktoren, eine der gegenwärtigen

Herausforderungen bei der Datenintegration für Cloud-Applikationen. Darüber hinaus bestehen zahlreiche kleine und große Herausforderungen, von denen nachfolgend fünf hervorgehoben sind, die für eine besondere Verschärfung im Cloud-Umfeld sorgen:

- **Abhängigkeit vom Schnittstellenumfang**  
In aller Regel erlaubt eine echte SaaS-Applikation keinen direkten Zugriff auf die Datenbanken, die der Applikation zugrunde liegen. Der Anbieter gewährt Zugriff auf die Daten in der Applikation ausschließlich über Schnittstellen mit Soap-Web-Services, REST, CSV-Dateien etc. Somit gilt, dass alles, was nicht durch diese Schnittstellen passt, nicht exportiert oder importiert werden kann – zumindest nicht mit Standardmitteln. Kandidaten hierfür sind je nach System Dateianhänge, Zeitstempel für Erzeugung und Aktualisierung von Datensätzen, Reports, Sichtbarkeitsregeln, Zugriffsrechte, Custom-Metadaten etc. Bei Dateianhängen und Zugriffsrechten hat Oracle CRM On Demand über die letzten Releases Fortschritte gemacht. Mit den sonstigen Aspekten können die OCOD Schnittstellen gut umgehen – bis auf Reports. Deren Ex- und Import wird noch immer nicht sauber unterstützt.
- **Strikte Vorgaben beim Datenformat**  
Feste Vorgaben für die Formatierung von Datenfeldern – korrektes E-Mail-Format für E-Mail-Adressen, vordefinierte Picklists, Einmaligkeit von Feldeinträgen bei Schlüsseln, Beschränkungen der Zeichenanzahl etc. – werden von den Anbietern bzw. in der Konfiguration bewusst vorgegeben, um die Datenqualität im laufenden Betrieb hochzuhalten. Hier zeigt die Erfahrung, dass Cloud-Applikationen nahezu immer strikter sind als traditionelle Lösungen, da für sie wie oben erwähnt Schnittstellen bewusst definiert sowie implementiert werden müssen und Speicherplatz wie auch Datenstrukturen für einen kosteneffizienten Betrieb beim Anbieter möglichst schlank gestaltet werden. Eine Herausforderung stellen die strikten Vorgaben dar, wenn die Daten, die aus einem anderen System stammen und importiert werden sollen, im Ausgangsformat nicht diesen Vorgaben entsprechen.

- **Beschränkung in der Datenübertragung**  
Aus wirtschaftlichen Gründen drosseln Anbieter von Cloud-Applikationen den Durchsatz auf Integrationschnittstellen meist noch unter die technischen Belastungsgrenzen ihrer Systeme. Oft gibt es sowohl Beschränkungen auf dem ein- und ausgehenden Datenstrom als auch auf den Rechenkapazitäten, die zur Verarbeitung von einkommenden Daten und zur Aufbereitung von ausgehenden Daten notwendig sind. Das betrifft bei den Web Service Allotments von Oracle konkret sowohl die Bandbreite für die rohen Übertragungsdaten als auch die Anzahl der Operationen, die in der SaaS-Software für das Einfügen, Löschen, Modifizieren oder Ausliefern von Daten notwendig sind.
- **Fremdbestimmte Verfügbarkeit**  
Die Verfügbarkeit einer Cloud-Applikation unterliegt weder der Kontrolle des Kunden noch des Dienstleisters für die Datenintegration. Wartungsfenster können von diesen Parteien nicht festgelegt werden, sondern werden vom Betreiber der Cloud-Applikation bestimmt und nach seiner Kommunikationspolitik verkündet. Eine rechtzeitige Ankündigung oder die Berücksichtigung der Zeitzone des Kunden ist nicht gewährleistet. Durch kurzfristig oder gar nicht angekündigte Ausfälle können Modifikationen auf dem Datenbestand besonders bei einer automatischen Synchronisation verloren gehen. Das wiederum kann zur Folge haben, dass die Konsistenz im Datenbestand verletzt wird.
- **Keine Transaktionssicherheit**  
Bei Soap und REST APIs beschränken sich die einzelnen Aufrufe auf kleine Operationen. Schon das Einfügen einer Anzahl von mehr als zwanzig Einträgen für eine Entität muss bei der Web-Service-API von OCOD auf mehrere Aufrufe verteilt werden. Verschiedene Operationen mit sequenzieller Abhängigkeit können mit den Bordmitteln der Web Services nicht transaktionssicher verbunden werden.

### Qualitätsmerkmale

Planbarkeit, Wartbarkeit, Kosteneffizienz und Geschwindigkeit sind wichtige Merkmale für eine gelungene Datenintegrationslösung. Da die Datenintegration aber immer

einem oder mehreren Systemen dient, die mit den Daten – und selbstverständlich ihren Nutzern – zur Wertschöpfung im Unternehmen beitragen, gibt es ein übergeordnetes Qualitätsmerkmal: Datenqualität.

Was kosten Daten von schlechter Qualität ein Unternehmen? Sie führen schnell zu Schäden, die weit über vermeintlichen Ersparnissen einer unausgegorenen Datenintegrationslösung liegen. Ohne Datenqualität verliert jedes andere genannte Merkmal seine Bedeutung und die gesamte Lösung ihren Wert.

### Werkzeuge

Prinzipiell eignet sich eine Vielzahl von modernen Middleware-Angeboten zur Umsetzung einer Datenintegrationslösung für Cloud-Applikationen. Oracle SOA Suite und Oracle Data Integration Suite liefern hier vielfältige Produkte, aus denen sich eine solide Datenintegrationslösung bauen lässt. IBM WebSphere Cast Iron, Talend, Jitterbit und Dell Boomi als interessanter Cloud-Ansatz seien ebenfalls erwähnt. Aber auch mit vergleichsweise schlanken eigenen Software-Tools realisiert das Unternehmen des Autors umfangreiche Datenintegrationslösungen beispielsweise für Oracle CRM On Demand.

Nach Einarbeitung in eine geeignete Middleware lassen sich eine saubere Anbindung an die Applikationen sowie eine gute Wartbarkeit realisieren. In jedem Fall aber gilt, dass nicht die eingesetzte Software an sich die Qualität einer Datenintegrationslösung gewährleistet. Der größte Faktor ist eine systematische Herangehensweise an die Entwicklung und Einführung einer Lösung und die Architektur der Umsetzung in der Middleware.

### Die Herausforderungen in der Cloud meistern

Unabhängig von der Wahl der Middleware können folgende Hinweise dabei helfen, die aufgeführten fünf Herausforderungen zu meistern:

- Um der festen Vorgabe des Schnittstellenumfangs erfolgreich zu begegnen, muss schon vor der Umsetzung einer Datenintegrationslösung das Zielsystem analysiert werden. Dabei ist das Studieren der Dokumentation zur entsprechenden API unerlässlich. Zudem sind

die benötigten API-Aufrufe zu testen. Für OCOD steht die API-Dokumentation der Web Services im Oracle Web Services On Demand Guide zum entsprechenden Release. Aber auch bei diesem API gilt es, vorher die Aufrufe zu testen, da verschiedene Einschränkungen nicht dokumentiert sind und einige wenige Felder aus den WSDLs sich nicht verhalten, wie es die Dokumentation vermuten lässt. Wenn eine einzige Schnittstelle nicht ausreicht, um alle notwendigen Daten zu behandeln, hilft meist die Kombination von Schnittstellen. Das reicht manches Mal bis zur Extraktion und zum Import direkt über die GUI einer Applikation.

- Strikte Vorgaben beim Datenformat setzen für den erfolgreichen Aufbau einer Integrationslösung ebenfalls gute Kenntnisse der festen und konfigurierbaren Beschränkungen voraus. Das Studium der oben erwähnten Dokumentationen bildet auch hierbei eine gute Grundlage. Zusätzlich sollte von Beginn an eine eigene Dokumentation für die individuelle Konfiguration der Schnittstellen einer Applikation erstellt und gepflegt werden. Außerdem gilt es, für die korrekte Datenformatierung die Quelldaten zu analysieren und Regeln zu definieren für die Transformation in ein Format, das mit den Vorgaben des Zielsystems kompatibel ist.
- Ist eine Beschränkung in der Datenübertragung zu erwarten, sollte die Logik auf Seiten der Middleware von Beginn an darauf ausgelegt sein. Auch wenn bei einer Synchronisation das initiale Volumen im Rahmen der Beschränkungen noch problemlos bewältigt werden kann, ist meist davon auszugehen, dass das Volumen beim möglichen Ausbau der Integrationslösung und beim Gedeihen des Unternehmens wächst. Die entsprechende Optimierung der Logik wird durch manche Middleware zumindest in Teilen unterstützt. Bei komplexeren Integrationslösungen bedarf sie aber immer auch eines darauf ausgerichteten Architekturaufbaus. Ziel aller Bemühungen ist letztendlich die Reduktion auf möglichst wenige und möglichst effiziente Operationen.
- Die fremdbestimmte Verfügbarkeit einer Cloud-Applikation sollte mindes-

tens eine ausführliche Protokollierung von Ausfällen und Fehlern auf Seiten der Middleware zur Folge haben. Warnungen sollten aktiv an verantwortliche Personen gesendet werden und Verweise auf weiterführende Informationen enthalten. Hierbei kann schon ein einfacher Monitoring-Mechanismus in Kombination mit aussagekräftigen Log-Files wertvolle Dienste leisten. Darüber hinaus kann die Middleware einen wesentlichen Teil zur Stabilisierung der Datenintegration bei Ausfällen in der Verfügbarkeit leisten, wenn sie Fehler während der Laufzeit interpretiert und je nach Fehlerfall automatisch eine einfache Wiederholung der fehlgeschlagenen Aufrufe vornimmt oder gar eine ausgeklügelte Korrekturmaßnahme einleitet.

- Transaktionssicherheit kann in Grenzen in einer geeigneten Middleware realisiert werden. Dabei hilft bereits die oben beschriebene automatische Behandlung fehlgeschlagener Aufrufe. Weitere wesentliche Voraussetzungen sind die Erkennung des Transaktionsstatus sowie Rollback- und Korrekturmechanismen in der Middleware. Sorgfalt beim Design von Prozessen in der Integrationslösung ist zudem unerlässlich. Wichtig ist jedoch auch das Bewusstsein, dass Transaktionssicherheit bei einer komplexeren Datenintegration für die Cloud kaum für jeden Fehlerfall gewährleistet werden kann. Daher sind ein umfassendes Logging in der Integrationslösung und dokumentierte Hinweise zur Behandlung von Fehlerfällen notwendig.

### Fazit

Cloud Computing und SaaS bringen neue Herausforderungen bei der Datenintegration. Zeitgemäße Middleware bietet ein gutes Handwerkszeug für die Realisierung von Datenintegrationslösungen und löst etliche der Herausforderungen. Mit einer systematischen Herangehensweise und notwendiger Umsicht lassen sich schon mit leichtgewichtigen technischen Mitteln kleine und mittlere Integrationslösungen hervorragend meistern. Die hier vorgestellten Hinweise helfen dabei, auch im Cloud-Umfeld zu stabilen Lösungen zu kommen.

*Dr. Daniel Rolli  
rolli@conemis.com*