

Sofortkopien – Minutenschnell in Selbstbedienung

Karsten Stöhr
Delphix
Frankfurt

Schlüsselworte

Virtual Data, Datenkopien, Instant Cloning, Self Provisioning, Live Archive, Thin Provisioning, DevOps.

Einleitung

Ob die Entwicklung eine neue Umgebung benötigt, ob ein QA-Test mit dem Gesamtdatenbestand ausgeführt werden soll, ob die Revision nochmal den Stand vor 6 Monaten prüfen will, egal für welchen Zweck, alle Projekte müssen auf die Einrichtung einer entsprechenden Umgebung warten. Während auf breiter Front die Server-Virtualisierung für eine schnelle und platzsparende Bereitstellung der Plattform genutzt wird, verschlingt die Bereitstellung der Daten immer noch viel Zeit. Oder es leidet die Qualität des Projekts, weil zur Vereinfachung eine andere Kopie wiederverwendet oder mit einem kleineren Testdatenbestand gearbeitet wird.

Dieser Vortrag zeigt wie jedem Projekt und jedem Team für seine Zwecke seine passenden virtuellen Daten zur Verfügung gestellt werden können - in Minutenschnelle, ohne dabei den mehrfachen Platz der Daten zu verbrauchen. Dabei wird ein intensiver Blick auf die Agile Data Platform von Delphix geworfen.

Eine Sisyphos-Arbeit: Daten kopieren

Datenbank-Administratoren sichern nicht nur die Verfügbarkeit der Daten, auf die aktuell von den Unternehmens-Anwendungen zugegriffen wird. Sie müssen auch den künftigen Bedarf im Blick haben und Strategien und Lösungen für Datenwachstum, Leistung der Infrastruktur und Einführung neuer Technologien entwickeln. Doch anstatt sich um strategische Aufgaben kümmern zu können, müssen DBAs in den meisten Organisationen einen großen Teil ihrer Zeit – bis zu 50% - für stupide Tätigkeiten opfern, nämlich das Kopieren und Bewegen von Daten sowie deren Richtigkeit zu überprüfen.

Fakt ist, wir kopieren Daten – immer und immer wieder.

Viele Administratoren versuchen, die sich ständig wiederholenden, manuellen Tätigkeiten mit Hilfe von selbst geschriebenen Skripten zu automatisieren. Trotzdem müssen sie aufgrund der Komplexität von Anwendungsprojekten immer wieder individuelle Anfragen nach Aktualisierung, Wiederherstellung und Integration der Daten manuell bearbeiten.

Ein Ausweg: Datenvirtualisierung

Die Bereitstellung von Servern war in der Vergangenheit ähnlich zeit- und kostenintensiv wie heute die Bereitstellung von Daten häufig noch ist. Eine deutliche Erleichterung in der Bereitstellung von Servern brachte der Wechsel von physischer Hardware zu virtuellen Maschinen. Ein sogenannter Hypervisor läßt einen einzelnen physischen Server aussehen wie mehrere separate Server durch eine intelligente gemeinsame Nutzung der vorhandenen Hardware-Ressourcen wie CPU und Speicher. Eine virtuelle Maschine bereitzustellen erfordert nur Minuten im Gegensatz zum mehrtägigen Aufwand, einen neuen physischen Server einzurichten.

Auf ähnliche Weise läßt sich auch die Bereitstellung von Daten beschleunigen. Wenn wir eine neue Kopie einer Datenbank - z.B. für Entwicklung oder Test - erstellen, dann sind zunächst alle Daten der Kopie identisch zur ursprünglichen DB. Erst im Laufe der Bearbeitung werden gewisse Daten

geändert oder neue erzeugt. Aber der Großteil der Daten bleibt identisch. Dies ermöglicht doch, geeignete Technologie vorausgesetzt, daß dieselben Datenblöcke für mehrere Kopien gemeinsam verwendet werden können, neudeutsch: sharing – ähnlich wie ein Hypervisor die Hardware-Ressourcen im Sharing für mehrere virtuelle Server verwendet.

Nur die von der jeweiligen Instanz individuell geänderten Datenblöcke müssen getrennt gespeichert werden (copy-on-write). Die Kombination aus shared Datenblöcken und individuell geänderten Datenblöcken läßt mehrere Kopien derselben Datenbank jeweils als eine vollständige separate Instanz erscheinen. Dabei verschlingt jede weitere Instanz nur einen Bruchteil des ursprünglichen Plattenplatzes, nämlich die Größe der individuell geänderten Daten – die bei der Instanzierung einer neuen Kopie ja noch gleich Null ist. Auf diese Weise dauert die Erstellung einer neuen Instanz selbst bei sehr großen Datenbanken nur noch Minuten statt diverser Stunden bis Tage.

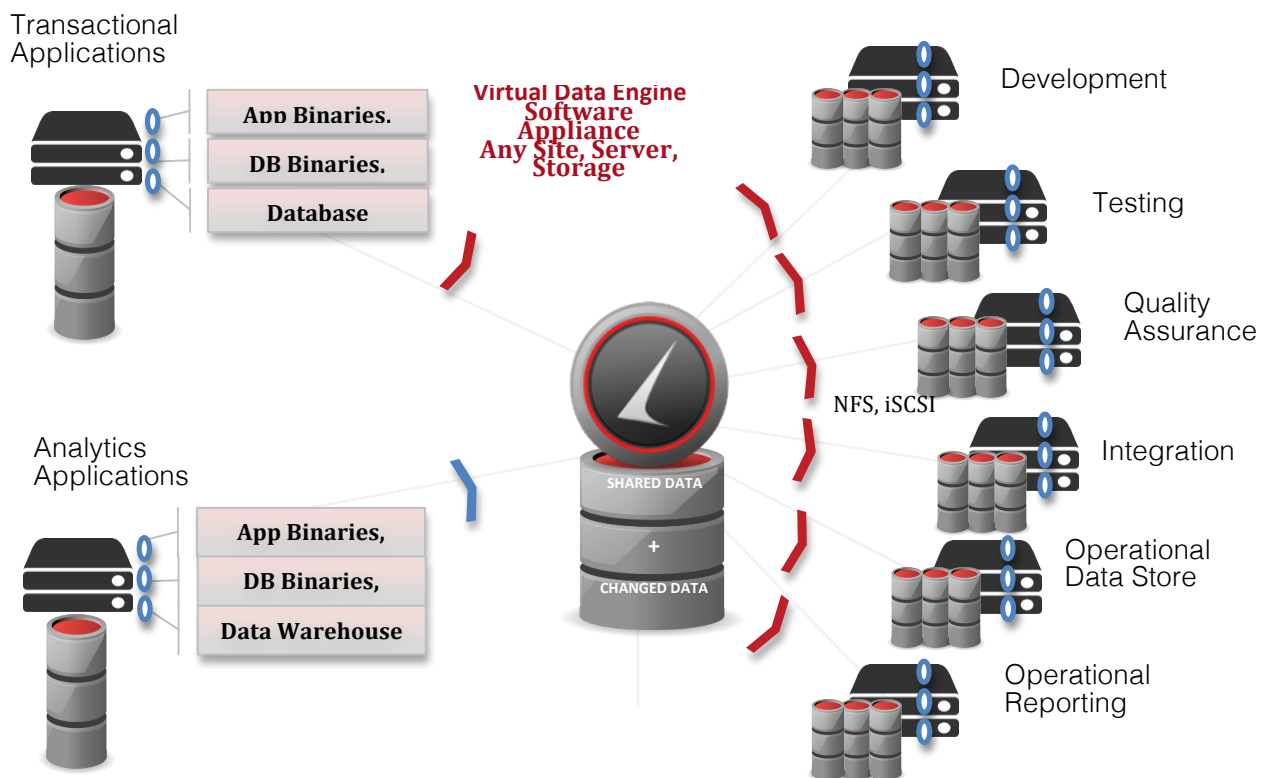


Abb. 1: Virtuelle Datenplattform

Zu den Applikationsdaten gehören nicht nur in Datenbanken gespeicherte Daten sondern auch diverse Dateien. Eine geeignete Plattform zur Datenvirtualisierung muß daher auch die entsprechenden Dateien auf ähnliche Weise bereitstellen können.

Zweiter Schritt: Selbstbedienung

Anwender, Entwickler, Tester stellen ihre Anfragen nach Datenumgebungen an die Datenmanagementabteilung. Dort werden die Anfragen geprüft, priorisiert und ausgeführt. Mittels einer Datenvirtualisierung wie oben geschildert können Administratoren diese Anfragen schneller und zuverlässiger beantworten, doch sind DBAs häufig stark belastet und Anfragen mit höherer Priorität schieben andere Anfragen nach hinten.

Nehmen wir an, ein Entwickler möchte eine neue Funktion mit großem Datenbestand testen, um die Performanz zu messen und eventuell zu verbessern. Dazu muß er denselben Test mehrere Male mit verschiedenen Parametern ausführen, und vor jedem neuen Testlauf müssen die Daten wieder auf den originären Stand zurückgesetzt werden. Der Testlauf selbst dauert vielleicht weniger als eine Stunde, aber für das Zurücksetzen der Daten muß der Anwender jedes Mal wieder eine Anfrage an die Administration stellen und auf deren Ausführung warten.

Wenn für jede Datenmanagementoperation ein Anwender sich erst an die IT Administration wenden muß, ist das ineffektiv und kostet zu viel Zeit. Die gesamte Projektlaufzeit wird verlangsamt. Wünschenswert ist es daher, daß einfache Operationen auf vordefinierten Umgebungen wie z.B. das Zurücksetzen (Reset) oder Aktualisieren (Refresh) der Daten vom Anwender selbst ausgelöst werden kann. Dies führt zu einer Aufgabenteilung zwischen der IT Infrastruktur und dem Management der Daten aus Anwendersicht. Die IT Administration behält volle Kontrolle über die Ressourcen und deren Verteilung, und kann eine definierte Datenumgebung (ein Template) einem Anwender oder einer Gruppe zuordnen. Die Anwender haben dann die Möglichkeit, in Selbstbedienung die zugewiesenen Umgebungen zu nutzen, zu aktualisieren, zurück zu setzen oder mit anderen Anwendern zu teilen – ohne die physischen Ressourcen zu kennen, die ihre Daten bereitstellen.

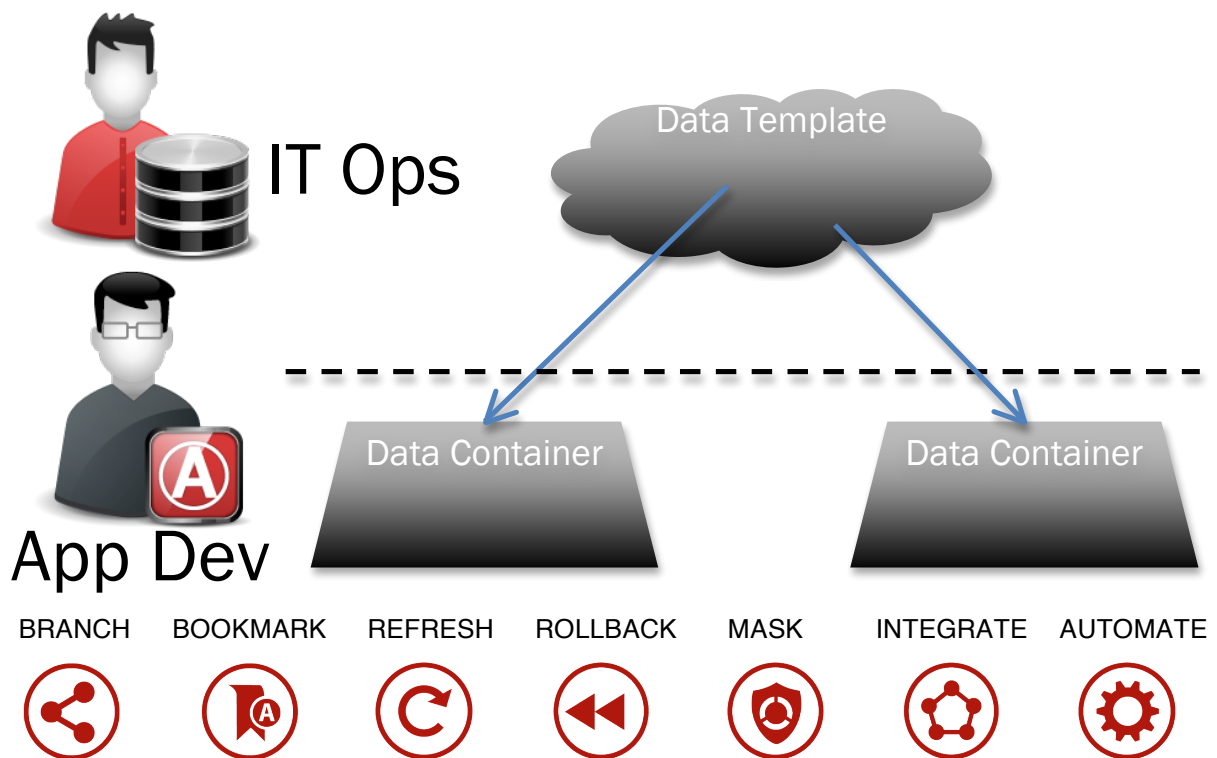


Abb. 2: Selbstbedienung und Trennung der Aufgaben von IT Administration und Anwendern

Delphix Agile Data Platform

Eine Plattform zur Datenvirtualisierung, welche alle zuvor geschilderten Anforderungen erfüllt ist die Agile Data Platform von Delphix, einem Anbieter, der letztes Jahr von den Mitgliedern der britischen Oracle User Group zum Technologiepartner des Jahres gewählt wurde. Grund genug, diese Software einmal näher zu betrachten. Tatsächlich hat mich das Produkt so sehr überzeugt, daß ich nach dessen Evaluierung dem Unternehmen Delphix beigetreten bin.

Funktionsweise

Die Delphix Engine ist eine Applikation mit eigenständiger operativer Umgebung, welche als virtuelle Appliance geliefert wird. Delphix verwendet Standard-APIs, um sich mit physischen Quelldatenbanken zu verbinden, ihre Datenblöcke zu kopieren und eine „goldene Kopie“ zu erzeugen, die sich 20 und mehr virtuelle Umgebungen teilen können. Dabei nutzt die Delphix Engine intelligente Filterung und Komprimierung, um die Kopie einer Quelldatenbank auf bis zu 50% ihrer originalen Größe zu reduzieren. Die Kopie der Quelldatenbank wird in der Delphix Engine zusammen mit allen inkrementellen Updates gespeichert als sogenannte dSource nach Delphix Terminologie.

Fortwährende Synchronisation

Nach dem initialen Laden hält die Delphix Engine die Synchronisation mit der Quelle aufrecht basierend auf konfigurierbaren Strategien, zum Beispiel einmal täglich oder innerhalb von Sekunden nach der jeweils letzten Transaktion. Einmal mit ihr verbunden führt Delphix einen zeitlichen Ablauf - Timeflow – der Quelldatenbank. Diese fortlaufende Aufzeichnung der Datei- und Logänderungen wird über ebenfalls konfigurierbare Regeln aufbewahrt, z.B. "hebe die letzten zwei Monate auf".

Automatisierte Bereitstellung

Von jedem beliebigen Zeitpunkt innerhalb des Aufbewahrungsfensters kann augenblicklich eine virtuelle Datenbank (in Delphix Terminologie eine VDB) von der Delphix Engine bereitgestellt werden. VDBs werden vom selben Speicherplatz des Timeflows der dSource beliefert (shared storage), verbrauchen also keinen zusätzlichen Plattenplatz.

Virtuelle Daten

Mehrere VDBs können von beliebigen verschiedenen Zeitpunkten innerhalb des Timeflows sekundengenau erzeugt werden. Eine VDB kann also ein „live“ Abbild des jetzt aktuellen Stands der Daten enthalten, eine andere dagegen einen historischen Stand von Bedeutung, z.B. den Stand einen Tag vor Entdeckung eines Fehlers, eine Woche vor einem Systemupgrade oder die Daten vor einem Jahr auf Anfrage eines Compliance-Verfahrens.

Einmal erzeugt, ist jede VDB eine eigenständige und unabhängige Datenbank mit vollständigen Lese- und Schreibzugriffen. Änderungen an der VDB durch Benutzer oder Applikationen werden in neue, komprimierte Blöcke im Delphix Storage geschrieben. Von einer VDB können wiederum neue VDBs erzeugt werden, und die Daten innerhalb einer VDB können von der dSource oder der übergeordneten VDB neu geladen werden (refresh).

Applikations Live-Archiv

Neben Daten aus Datenbanken und Dateien kann Delphix auch Kopien der Binärdateien von Applikationen erstellen und aufbewahren. Mit allen drei Ressourcen kann Delphix ein Live-Archiv von kompletten Applikationen führen, welches eine Wiederherstellung der gesamten Applikation von einem beliebigen früheren Zeitpunkt ermöglicht, inklusive der Applikation, ihrer Konfiguration und ihrer Daten zu dem Zeitpunkt.

Minutenschnell in Selbstbedienung

Delphix bietet eine einfach zu bedienende Web-Oberfläche, welche es allen Anwendern aus Entwicklung, Qualitätskontrolle, Analyses, etc. ermöglicht, aus den ihnen zugewiesenen Ressourcen ihre eigenen virtuellen Datenumgebungen zu erzeugen und aus Datensicht zu managen – ohne immer erst eine Anfrage an die IT-Administration stellen zu müssen.

RESTful API

Alle Funktionen der Delphix Engine können über ein dokumentiertes API Set in bestehende Werkzeuge und Routinen zur automatischen Provisionierung von Anwendungsumgebungen integriert werden.

Erweiterungen

Compliance Engine – Live Archive und Data Masking

Modernization Engine – Unix to Linux Transformation

Cloud – Delphix for Amazon Web Services und Delphix for VMWare vCloud Air

SAP App Data – SAP spezifisches Toolkit

EBS App Data – Oracle E-Business Suite spezifisches Toolkit

Kontaktadresse:

Karsten Stöhr

DELPHIX

THE SQUAIRE 12

D-60549 Frankfurt

Telefon: +49 (69) 959325 383
E-Mail karsten.stoehr@delphix.com
Internet: www.delphix.com