

Whitepaper

Externe Speicherung von Binary Large Objects (BLOBs) mit SharePoint 2007 sowie SQL Server 2005 / 2008

Andreas Glaser, Michael Pertek
namics ag
23. September 2008

Inhaltsverzeichnis

1	Einleitung	3
2	Unstrukturierte Daten	4
2.1	Business value	4
2.2	SharePoint 2007	5
3	External BLOB Storage	7
3.1	Limitierungen und Abwägungen	8
3.2	Implementation	9
4	FileStream	10
5	Zusammenfassung	11
6	Anhang	12
6.1	Abbildungsverzeichnis	12
6.2	Kontakt	12

1 Einleitung

Dieses Whitepaper zeigt Möglichkeiten und Restriktionen bei der Speicherung von Daten mittels SharePoint 2007 und dem SQL Server 2005 (oder dem SQL Server 2008) extern auf dem Dateisystem zum jetzigen Zeitpunkt auf.

Schlagwörter wie „FileStream“, „EBS Provider“ und „externe Speicherung von grossen Datenmengen“ kursieren innerhalb der Anwendungsbereiche von SharePoint und SQL Server und werden viel diskutiert. Hinsichtlich der gegebenen Verwendungsmöglichkeiten dieser Features und den nicht zu unterschätzenden Einschränkungen gibt es allerdings aktuell kaum dokumentierte Erfahrungen im Einsatz dieser Features. Diese Zusammenstellung der zentralen Fakten und Lösungsvarianten beleuchtet den richtigen Umgang mit grossen Datenmengen und deren Speicherung.

Das Whitepaper vermittelt Entscheidungsträgern die Problematik des Umgangs mit grossen Datenmengen in SharePoint und stellt die gegebenen Lösungsmöglichkeiten gegenüber.

2 Unstrukturierte Daten

Grundlage dieser Ausführungen ist die Annahme, dass aufgrund des starken, fast explosionsartig **ansteigenden Aufkommens** an digitalen Daten und der **einhergehenden Reduktion** der Anschaffungs- und Betriebskosten von Hardwarekomponenten zur Speicherung des vermehrten Datenaufkommens viele Firmen unstrukturierte Daten in einer Datenbank speichern. Unter unstrukturierten Daten, auch **Binary Large Objects (BLOB)** genannt, versteht man Dokumente, digitale Bilder, Audio oder Videoformate. Immer häufiger werden diese Daten zu einem geschäftskritischen Bestandteil in Firmen und ihre Vorhaltung gewinnt an strategischer Bedeutung.

Ein Spital beispielsweise, speichert Röntgenbilder in einer Datenbank. Sowohl für den Patienten als auch für den Arzt sind der einfache und schnelle Zugriff, sowie die hohe Verfügbarkeit dieser Bilder von enormer Bedeutung und im wahrsten Sinne des Wortes „überlebenswichtig“.

2.1 Business value

Die Speicherung von strukturierten und unstrukturierten Daten geschieht im SQL Server 2005 auf dieselbe Art und Weise. Die erste Darstellung im linken Bild zeigt, wie sowohl Dateien als auch Metadaten dieser Dateien in einer relationalen Datenbank gespeichert werden. In der rechten Darstellung wird derselbe Vorgang für unstrukturierte Daten innerhalb des SQL Servers 2008 abgebildet.

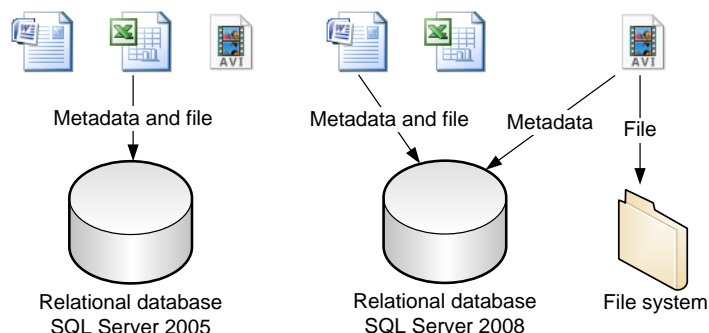


Abbildung 1: Unstrukturierte Daten im SQL Server 2005 und 2008

Der Zugriff auf BLOB's, welche in der Datenbank gespeichert sind, ist allerdings in der Regel zu wenig performant. Vor allem bei grossen Dateien wie Videos ist der Unterschied aus zeitlicher Perspektive (Reaktions- und Darstellungszeit) erheblich verglichen mit dem Zugriff auf ein Video, welches auf dem Dateisystem gespeichert wurde.

Der SQL Server 2008 dagegen kann BLOB's auch direkt auf dem Dateisystem speichern und die dazugehörigen Metadaten in einer Datenbank ablegen. Die Performance ist dadurch deutlich höher und für den Benutzer spürbar. Zusätzlich ist die Grösse der gespeicherten Dateien nicht wie beim SQL Server 2005 auf 2GB beschränkt.

Der SQL Server 2008 bietet hinsichtlich Wartbarkeit, Verfügbarkeit, Sicherheit und der Möglichkeit, grosse Daten extern zu speichern (Stichwort **FileStream**) einen erheblichen Mehrwert im Vergleich zum seinem Vorgänger.

Der SQL Server 2005 bietet nicht die Möglichkeit der externen Speicherung von Daten wie oben beschrieben. Zusammen mit SharePoint 2007 ist es dennoch möglich, die externe Speicherung über Sharepoint zu realisieren (Stichwort **External BLOB Storage**). Die genaue Vorgehensweise wird nachfolgend beschrieben.

2.2 SharePoint 2007

Nachdem der SQL Server 2008 der Öffentlichkeit zur Verfügung gestellt wurde, kündigte der *Microsoft SharePoint Team Blog*¹ die Unterstützung des SQL Server 2008 durch „SharePoint Products and Technologies“ an. Dies bedeutet, dass zukünftig der SQL Server 2008 als Datenbankmanagement-System mit *Microsoft Office SharePoint Server 2007 (MOSS 2007)* und *Windows SharePoint Services 3.0 (WSS 3.0)* eingesetzt werden kann.

Unter der Voraussetzung, dass **Service Pack 1** installiert wurde, ergeben sich daraus zwei Lösungsansätze, um Daten ausserhalb des SQL Server zu speichern:

- » **External BLOB Storage – EBS (SQL Server 2005)**

¹ <http://blogs.msdn.com/sharepoint/default.aspx>

Whitepaper	Externe Speicherung von Binary Large Objects (BLOBs) mit SharePoint 2007 sowie SQL Server 2005 / 2008 namics ag	23. September 2008	Seite 6
------------	--	--------------------	---------

» **FileStream** (SQL Server 2008)

Der erste Lösungsansatz (EBS) ist mit SharePoint umzusetzen und unterstützt den SQL Server 2005. Der zweite Lösungsansatz (FileStream) ist ein Feature des neuen SQL Server 2008.

Die Lösungsansätze werden in Kapitel 3 kurz vorgestellt und auf ihre Funktionsweise hin bewertet.

3 External BLOB Storage

Mit dem *Windows SharePoint Services 3.0 hotfix package* vom 25. Mai 2007² hat Microsoft eine Schnittstelle (API) zur Verfügung gestellt, um Daten ausserhalb des SQL Server zu speichern. Dieser Hotfix wurde später in das Service Pack 1³ integriert, zusätzlich gibt es eine technische Dokumentation zur Nutzung dieser API⁴, die Funktionsweise wird anhand untenstehender Abbildung verdeutlicht.

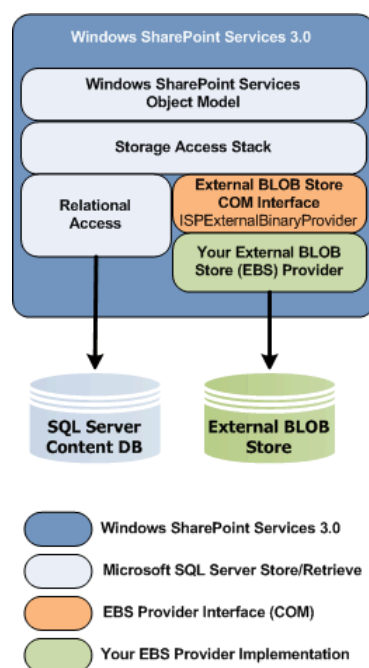


Abbildung 2: External BLOB Storage Architektur⁵

Um Daten mit den WSS 3.0 extern zu speichern, muss ein separater EBS Provider implementiert werden. Die dazugehörige Schnittstelle ist Bestandteil der WSS 3.0 inklusive SP1 und daher unabhängig vom genutzten Datenbankmanagementsystems.

² <http://support.microsoft.com/kb/937901/en-us>

³ http://kyletillman.net/blog/post/SharePoint_External_Binary_Store-Part1.aspx

⁴ <http://msdn.microsoft.com/en-us/library/bb802976.aspx>

⁵ <http://msdn.microsoft.com/en-us/library/bb862195.aspx>

Sowohl die strukturierten Daten als auch die Metadaten von BLOB's werden weiterhin in der Datenbank gespeichert. Die BLOB's werden allerdings auf dem Dateisystem gespeichert und mittels einer ID referenziert.

3.1 Limitierungen und Abwägungen⁶

Die Umsetzung eines separaten **EBS Provider** hat wesentliche Auswirkungen auf eine Sharepoint Farm. Sowohl Backup und Restore als auch Garbage Collection sind hiervon betroffen. Aufgrund des einfachen Designs der Schnittstelle sollten folgende Limitierungen beachtet werden und zusätzlich der entstehende Aufwand dem zu erwartenden Nutzen sorgfältig gegenübergestellt werden.

- » **Kein Update von extern gespeicherten Daten:** Jedes Editieren einer extern gespeicherten Datei führt zur Erstellung einer Kopie dieser Datei, wodurch der Bedarf an Speicherplatz bei sehr grossen Dateien (> 1 GB) enorm wächst. Ein eigener Garbage Collector zum Löschen alter Dateien sollte zwingend implementiert werden.
- » **Migration:** In der Datenbank existierende Dateien können entweder als Variante 1 per Site Level Backup und Restore aus der Datenbank auf das Dateisystem migriert werden.
Als Variante 2 könnten andererseits alle Dateien aktualisiert werden, was dazu führt, dass diese auf dem Dateisystem gespeichert werden und die alte Version aus der Datenbank gelöscht wird.
- » **Backup und Restore:** Die Backup und Restore Funktionalitäten der WSS 3.0 berücksichtigen extern gespeicherte Dateien leider nicht. Für diesen Fall muss ein separater Disaster Recovery Plan erstellt werden.
- » **Farm-Level:** In der aktuellen Version kann die externe Speicherung von Daten über EBS nur auf Farm Level aktiviert werden, hiervon sind **alle** Websites betroffen. Zusätzlich muss die Aktivierung auf allen Front-End-Webservern vorgenommen werden, da durch die Load Balancing Funktion der Server verschiedene Webserver angesprochen werden können.

⁶ <http://msdn.microsoft.com/en-us/library/bb862135.aspx>

- » **Performance:** Fehler im Speichervorgang oder mangelhafte Performance, welche durch den EBS Provider erzeugt werden, wirken sich direkt auf das Verhalten der SharePoint Seiten aus.
- » **Inkompatibilität:** In der folgenden (noch nicht veröffentlichten) Version von SharePoint (Full Release) wird die Schnittstelle syntaktisch anders sein. Daher kann nicht davon ausgegangen werden, dass die Implementierung mit zukünftigen Versionen ohne manuelle Eingriffe funktioniert.

3.2 Implementation

Sollen Daten mittels External BLOB Storage extern gespeichert werden, muss

- » eine COM Komponente basierend auf dem ISPEXternalBinaryProvider Interface
und
- » ein Garbage Collector

implementiert werden. Erfahrungen⁷ zeigen, dass dieser Aufwand hinsichtlich Komplexität und Testressourcen hoch ist. Eine ausführliche und erklärende Dokumentation zur tatsächlichen Implementierung liegt seitens Hersteller noch nicht vor.

⁷ <http://kyletillman.net/blog/?tag=/ispexternalbinaryprovider>

4 FileStream

Der SQL Server 2008 stellt ein neues Feature namens **FileStream** zur Verfügung, um die oben beschriebenen BLOB's extern zu speichern. Dabei handelt es sich um ein singuläres Attribut, welches auf eine Spalte vom Typ *varbinary* gesetzt werden kann. Sowohl Backup als auch Restore des SQL Server 2008 berücksichtigen extern gespeicherten Daten.

Um das Feature FileStream auf einer Datenbank nutzen zu können, müssen folgende drei Schritte durchgeführt werden:

- » Aktivierung von FileStream auf der SQL Server Instanz
- » Erstellung einer Datenbank inklusive FileGroup und FileStreams
- » Erstellung einer Spalte mit FileStream-Unterstützung

Da SharePoint 2007 die Datenbanken automatisch erstellt, wäre es prinzipiell möglich, die FileStream-Unterstützung nach Erstellung selbst zu aktivieren. Seitens Microsoft wird die Änderung einer SharePoint Datenbank allerdings nicht supported. Daher ist von dieser Möglichkeit aus Gründen der Wartbarkeit und Release-Fähigkeit dringend abzuraten!

Es spricht allerdings nichts gegen den Einsatz des SQL Server 2008 zusammen mit SharePoint 2007, da der SQL Server 2008 Verbesserungen gegenüber dem SQL Server 2005 mit sich bringt. Das FileStream Feature kann aber nicht gemeinsam mit SharePoint 2007 genutzt werden.

Voraussetzung für die Zusammenarbeit von SharePoint 2007 und dem SQL Server 2008 ist das Service Pack 1.

5 Zusammenfassung

Die externe Speicherung von Daten mit dem SQL Server 2005 oder dem SQL Server 2008 und SharePoint 2007 ist zum Zeitpunkt der Veröffentlichung des Whitepapers nicht empfehlenswert.

Der Aufwand für die Umsetzung des **EBS** ist äusserst hoch. Eine mögliche Inkompatibilität mit zukünftigen Versionen von SharePoint ist mehr als wahrscheinlich.

Bei dem Versuch der Aktivierung des **FileStream** Attributes auf einer vorhandenen SharePoint Datenbank handelt es sich um einen Eingriff in die Datenbankstruktur von SharePoint 2007. In der Regel ist ein Support von Microsoft im Anschluss an diese Modifikationen nur sehr schwer möglich, da Datenbanken per se nicht geändert werden dürfen.

namics empfiehlt daher, auf das nächste Service Pack oder die folgende Version von SharePoint zu warten, um den vollen Funktionsumfang zu erhalten.

Abhängig von den Anforderungen an die Datenhaltung (u.a. Verfügbarkeit oder Sicherheit) kann allerdings der Einsatz eines SQL Server 2008 gegenüber eines SQL Server 2005 durchaus sinnvoll sein.

Um für die Zukunft gerüstet zu sein, empfiehlt sich der Einsatz des SQL Server 2008, da eine deutlich bessere Unterstützung der externen Datenspeicherung durch die Veröffentlichung zukünftiger Updates wahrscheinlich ist.

6 Anhang

6.1 Abbildungsverzeichnis

Abbildung 1: Unstrukturierte Daten im SQL Server 2005 und 2008 4

Abbildung 2: External BLOB Storage Architektur 7

6.2 Kontakt

namics ag
Teufenerstrasse 19
CH-9001 St. Gallen
t +41 [0] 71 228 67 77
f +41 [0] 71 228 67 88
<http://sharepoint.namics.com/>

Andreas Glaser, Software Engineer
andreas.glaser@namics.com

Michael Pertek, Business Unit Manager
michael.pertek@namics.com