# Speicher in den Wolken - Amazon S3

Praxisnaher Überblick über die Einsetzbarkeit und die Technik des Cloud-Services Amazon S3 anhand einer Beispielimplementierung

Barbara von Kalm

Fakultät für Informatik Hochschule Mannheim Paul-Wittsack-Strasse 10 68163 Mannheim 626027@stud.hs-mannheim.de

Zusammenfassung CloudComputing und AmazonWebServices gehören zu den wichtigsten Schlagworten der IT im Jahre 2009. Ein Webservice von Amazon ist der Amazon Simple Storage Service. Leider sind die eigentliche Technik des Services und der Anwendungsbereich weitgehend unbekannt. Die Ausarbeitung zu diesem Thema soll eine Quelle darstellen, mit deren Hilfe Unklarheiten zum Amazon S3-Service ausgeräumt werden sollen. Als Ergebnis erhält der Leser einen umfassenden Überblick über das Einsatzgebiet und die Technik des Amazon S3-Service. Für die Tauglichkeit des Einsatzes in bestehende Softwareanwendungen wird eine Beispielimplementierung des Zugriffs auf den Service in Java durchgeführt.

Die IT-Trends für das Jahr 2010 von Gartner sehen Cloud-Computing an erster Position[2]. Einer der ersten, die diesen Trend mitgestalten und schon jetzt ein großes Portfolio an Services in einer Cloud anbietet, ist Amazon. Der Cloud-Service, der hier näher betrachtet werden soll heißt "Amazon Simple Storage Service" (kurz: Amazon S3). Hierbei handelt es sich um das Anbieten von Speicherkapazitäten im Internet.

Ziel dieser Ausarbeitung ist es grundlegende Fragen zu beantworten, wie Was steckt hinter Amazon S3?, Wo liegen meine Daten?, Wie kommen meine Daten in die Cloud und wie kommen sie wieder heraus? Auch die Kosten-Frage und der Nutzen für Firmen ist interessant und soll hier Beachtung finden.

Zunächst werden die Grundlagen, die zum Verständnis der Arbeit notwendig sind in einem Grundlagenkapitel erläutert. In der eigentlichen Ausarbeitung des Themas soll an erster Stelle der Amazon S3 näher beleuchtet werden. Das nachfolgende Kapitel beschreibt die verschiedenen Zugriffsmöglichkeiten auf den Service und stellt eine kleine Anwendungslösung in Java vor. Die Kosten und die Vor- und Nachteile werden ebenfalls Bestandteil der Ausarbeitung sein. Zuletzt

soll der Leser entscheiden können, für wen Amazon S3 geeignet ist, und für wen nicht.

### 1 Allgemeines zu diesem Dokument

Dieses Dokument dient als Einstieg in das Thema "AmazonS3". Hierbei handelt es sich um eine Arbeit, die im Rahmen eines Seminars zum Thema Cloud-Computing entstanden ist. Details zum allgemeinen Thema Cloud-Computing wurden bereits durch andere Seminarteilnehmer oder den Seminarleiter erörtert, weshalb nähere Ausführungen zu diesem Thema hier unbehandelt bleiben.

## 2 Grundlagen

Bevor der S3 näher betrachtet werden kann, ist es zunächst wichtig einige Grundlagen kurz zu erläutern, damit ein Endruck davon entsteht, was Cloud-Computing bedeutet und was Speicher in einer Cloud ist.

#### 2.1 Cloud-Computing

Definition: In der Literatur findet man inzwischen eine Vielzahl von Definitionen von Cloud-Computing. Ein zentralen Punkt beschreiben sie aber fast alle, nämlich dass die Cloud dem Kunden Technologien zur Verfügung stellt, dann, wenn er sie braucht, so lange er sie braucht und so viel er sie braucht. Diese Technologien werden aufgeteilt in Platform as a Service (PaaS), Infrastructure as a Service (Iaas) und Software as a Service (Saas). Den S3 von Amazon kann man der Infrastructure zuordnen. Cloud-Computing basiert auf zwei zentralen Technologien, Virtualisierung und WebServices. Der Einsatz dieser Technologien sorgt für die signifikanten Eigenschaften eines Cloud-Services, nämlich die hohe Skalierbarkeit und der Zugriff von überall auf der Welt.

### 2.2 Amazon Web-Services

Amazon bietet seid dem Jahr 2006 eine Reihe von webbasierten Services. Alle zusammen werden als Amazon Web Services bezeichnet (AWS). Die verschiedenen Webservices decken eine Vielzahl von möglichen Cloud-Services ab:

- Amazon Elastic Cloud Compute (Amazon EC2) Stellt Rechenkapazitäten zur Verfügung
- Amazon Simple Storage Service (Amazon S3) stellt skalierbaren Speicher zur Verfügung
- Amazon Simple Queue Service (Amazon SQS)

- AmazonCloudFront Speicherung von Nachrichten, zum Transfer zwischen verschiedenen Computern
- Amazon SimpleDB

Mit verschiedene Service-Level-Agreements (kurz: SLA) garantiert Amazon eine gewisse Ausfallsicherheit, die dem Kunden das Gefühl geben kann, dass seine Daten immer zur Verfügung stehen.

#### 2.3 Speicher in einer Cloud

Beim Speichern von Daten in einer Cloud kommt es vor allem darauf an, unbegrenzt viele Daten speichern zu können, benötigt man mehr Speicher, so bekommt man mehr Speicher. Außerdem sollen alle möglichen Daten bzw. Objekte gespeichert werden können. Über dies hinaus ist einer der wichtigsten Punkte beim ablegen von Daten in einer Cloud, dass diese immer verfügbar sind und entsprechende Backuplösungen angeboten werden. Online-Speicher wird von immer mehr Firmen angeboten, doch bisher konnte sich kaum ein Anbieter gegen Amazon durchsetzen. Der Vollständigkeit halber sind einige nachfolgend aufgeführt:

- Amazon S3
- Mossos CloudFS Beta
- Memopal
- $EMC^2$

Weiterhin gibt es verschiedene Arten von Speicher in einer Cloud:

- Persistenten Speicher
- Flüchtiger Speicher
- Elastic Block Storage

Amazon bietet all diese Speicherarten in der Cloud an. Der persistenter Speicher wird durch den S3 abgedeckt und der Flüchtige Speicher durch EC2-Instanzen. Dem Elastic Block Storage ist nachfolgend ein Abschnitt gewidmet.

#### 2.4 EBS

Elastic Block Storage (kurz:EBS) ist die Art und Weise, wie Amazon seinen persistenten Speicher anbietet. EBS ist vergleichbar mit einem Storage Area Network (kurz:SAN), bei dem es darum geht Speicher, der an verschiedenen physikalischen Orten liegen, zu einem großen Blockspeicher zusammen zu fassen, um Zugriff auf den Speicher über einen zentralen Server zu gestalten, der wiederum den Zugriff in Form von Laufwerken anbietet (siehe nachfolgende Abbildung 1).

Mit Hilfe von EBS ist es möglich in der Cloud 1GB- bis 1TB-große Instanzen zu generieren und zu verwalten. EBS-Instanzen sind die Grundlage dafür



Abbildung 1. Speicherzusammenfassung in einem SAN

EC2-Instanzen anzulegen. Damit diese Instanzen auch nach dem Abschalten des System wieder zur Verfügung stehen bietet Amazon die Möglichkeit Snapshots von dem angelegten EBS zu machen und diese im S3 zu speichern, so lässt sich der Zustand eines Systems zu verschiedenen Zeitpunkten wiederherstellen.

#### 3 Amazon S3 in der Praxis

Nach den Grundlagen, die im vorangegangen Kapitel beschrieben sind, soll nun der Amazon S3 in seinen Einzelheiten betrachtet werden. Der Leser soll hier erfahren, was Amazon S3 überhaupt ist, wie man ihn nutzen kann, wer ihn nutzen kann bzw. sinnvoll nutzen kann und vor allem, wie man darauf zugreifen kann. Dazu werden zwei Lösungen vorgestellt, der Zugriff über ein Browser-Plugin und der Zugriff über eine Java API.

### 3.1 Was ist Amazon S3

Wie im Grundlagenkapitel kurz angedeutet, handelt es sich beim AmazonS3 um persistenten Speicher, der in einer Cloud angeboten wird. Es ist möglich Objekte verschiedenster Art, z.B. Multimediadateien oder PDF-Dokumente oder was auch immer elektronisch vorliegt, in beliebig großer Menge in der Cloud anzulegen. Zur Organisation der Objekte werden sogenannte Buckets erzeugt, in denen die Objekte gespeichert werden. Der Name eines Buckets muss Amazonweit eindeutig sein. Es ist nicht möglich Buckets ineinander zu verschachteln. (Buckets sind also nicht vergleichbar mit Ordnern in einem Dateisystem). Die verschiedenen Objekte in einem Bucket haben alle einen eindeutigen Schlüssel und sind auch darüber anzusprechen.

Was gibt es nun also für Entitäten im Amazon S3:

Objects - Objekte sind die Dinge, die gespeichert werden sollen. Daten sind auch über URLs abrufbar (dazu später mehr), weshalb nicht nur die eigentlichen Daten, sondern auch Metadaten gespeichert werden. Bei diesen Metadaten handelt es sich um Daten, die vom Browser gelesen werden können, wie z.B. den Contenttype. Diese Metadaten werden in Key-Value-Paaren gespeichert wie in der nachfolgenden Abbildung 2 dargestellt ist.

```
MetaData for 'object':

Key: ETag - Value: "d41d8cd98f00b204e9800998ecf8427e"

Key: Date - Value: Mon Nov 23 11:50:00 CET 2009

Key: Content-Length - Value: 0

Key: id-2 - Value: y1UJBWI7Agv905nSsaaWD0Z96Axafj7rAmgh2v2bIz5iadjF7FENb/Bg3/4MJJ0X

Key: request-id - Value: FE21F32B55320E2B

Key: Content-Type - Value: application/octet-stream
```

Abbildung 2. Metadaten eines S3-Objects

- Buckets Ein Bucket ist eine Art Container, in dem die einzelnen Objekte abgelegt werden. Der Name eines Buckets muss Amazonweit eindeutig sein. Einen Bucket mit dem Namen Einbuchten zu erstellen ist deshalb nicht ratsam. Außerdem muss bei dem Erstellen von Buckets darauf geachtet werden, dass keine Sonderzeichen wie Slawisches verwendet werden.
- Keys Jedes Objekt, dass in einem Bucket abgelegt ist hat einen eindeutigen Schlüssel. Ein Objekt wird eindeutig aufgerufen durch die Kombination zwischen Bucket und Key des Objekts.

Es gibt einige grundlegende Operationen, die man mit Amazon S3 durchführen kann:

- Einen Bucket erstellen, um Daten drin speichern zu können.
- Daten-Objekte hochladen und in einem Bucket ablegen.
- Daten-Objekte aus einem Bucket downloaden.
- Auflistung aller Objekte in einem Bucket.

#### 3.2 Sicherheit

Sicherheit ist ein sehr zentrales Thema, wenn man sich mit Speicher in der Cloud beschäftigt, denn die eigenen Daten werden irgendwo auf der Welt gespeichert und man muss sichergehen können, dass es keinen unbefugten Zugriff darauf gibt oder dass die Daten plötzlich nicht mehr vorhanden sind. Amazon hat dazu einen Sicherheitsprozess implementiert, der verschiedenen Askpekte zum Thema Datenschutz aufgreift [9]. Dabei spielen Themen, wie Vertrauen in den Anbieter, physikalische Sicherheit und Backup eine Rolle. Zum Thema Vertrauen stellt Amazon sich selbst als vertrauenswürdiger Partner da. Um dies auch zu belegen lässt sich Amazon regelmäßig SOX-zertifizieren und bietet außerdem Industriespezifische Zertifizierungen für seine Kunden an. Auch um einen eventuellen Verlust von Daten muss man sich keine Sorgen machen. Alle Daten werden in der Cloud-Wolke redundant an mehreren physikalischen Orten gesichert. Die Backups sind normaler Bestandteil eines Servicevertrages und müssen nicht extra bezahlt werden.

Zugriffskontrolle auf AmazonS3-Objekte Der Erzeuger eines Buckets kann die Zugriffe auf diesen und die enthaltenen Objekte individuell gestalten. In

Abbildung ?? sieht man, wie man die Zugriffe auf ein Objekt individuell festlegen kann. Beim Anlegen eines Objektes ist immer nur der private Zugriff auf das Objekt vorgesehen.



Abbildung 3. Zugriffssteuerung eines AmazonS3-Objekts

Amazon bietet außerdem die Möglichkeit über SSL-verschlüsselte Endpunkte auf den S3 zuzugreifen. Möchte man Daten von seinem EC2-Volume zu AmazonS3 verfrachten, ist dies auch über SSL möglich.

Dennoch ist unbedingt darauf hinzuweisen, dass sensible Daten unbedingt außerhalb der Amazon-Wolke verschlüsselt werden sollten und erst dann gespeichert werden sollten, um unerwünschten Mitlesern nicht die Chance zu geben unbefugt an Information zu geraten.

## 3.3 Zugriffsmöglichkeiten

Nachdem nun bekannt ist, was Buckets sind, und dass man auf seine Daten aufpassen muss, kommt nun der spannende Teil: Wie kommt man an seine Daten in den Wolken?

Nachdem man sich mit einer Kreditkarte bei Amazon angemeldet hat, gibt es verschiedene Möglichkeiten neue Buckets zu erstellen, Daten abzulegen oder sonstiges zu tun. Um einen ersten Eindruck zu bekommen, aber auch um manuell Daten zu verwalten empfiehlt sich das Firefox-Plugin Amazon S3 Firefox Organizer(S3Fox). Mit Hilfe dieses Plugins kann man Datenobjekte, wie mit einer FTP-Anwendung in seine zuvor angelegten Buckets schieben. Selbst die Steuerung der Zugriffskontrolle ist über dieses Plugin möglich. Einen ersten Eindruck kann man in der nachfolgenden Abbildung erhalten 4. Auf der

linken Seite sind die lokalen Dateien zu sehen, auf der rechten Seite sind die Buckets und darin enthaltenen Objekte im Service dargestellt. In der aktuellen Aufnahme sieht man, dass in dem geöffneten Bucket ein Objekt vorhanden ist.

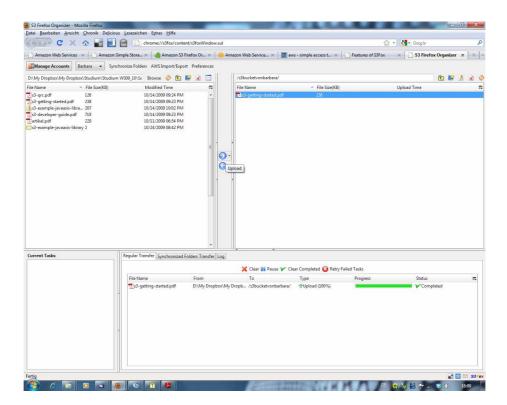


Abbildung 4. Amazon S3 Organizer Firefox Plugin

Um Amazon S3 in Anwendungen zu integrieren und somit Daten automatisiert verwalten zu können, bietet Amazon verschiedene Web-Services-Protokolle an. So ist es möglich über SOAP, REST oder Bittorrent auf den Service zuzugreifen. Es gibt bereits viele Bibliotheken für alle möglichen Programmiersprachen, die die grundlegenden Funktionen bereits implementiert haben und so den Zugriff auf den WebService so simple wie möglich gestalten. Für Java gibt es eine Bibliothek mit dem Namen jets3t, die in der nachfolgenden Beispielimplementierung zum Einsatz kommt. Grundsätzlich gibt es zwei sogenannte endpoints, um Amazon S3 per REST oder SOAP anzusprechen:

- http://s3.amazonaws.com
- oder SSL-verschlüsselt https://s3.amazonaws.com

Für die Benutzung mit SOAP hängt man hinten einfach /SOAP an. Nachfolgend sieht man ein Beispiel aus dem DeveloperGuide von Amazon 5. Hier der Header beim Zugriff per Get-Methode auf ein Bild dargestellt. Zu beachten ist die Authentifizierung, die immer nach diesem Muster abläuft: Authorization: AWS AWSAccessKeyId:Signature. Die APIs zu REST und SOAP sind bei Amazon umfassend beschrieben, weshalb sie hier nun nicht weiter beschrieben werden.

Request	StringToSign
GET /photos/puppy.jpg HTTP/1.1 Host: johnsmith.s3.amazonaws.com Date: Tue, 27 Mar 2007 19:36:42 +0000	GET\n \n \n
Authorization: AWS OPN5J17HBGZHT7JJ3X82: xXjDGYUmKxnwqr5KXNPGldn5LbA=	Tue, 27 Mar 2007 19:36:42 +0000\n /johnsmith/photos/puppy.jpg

Abbildung 5. Authentifizierung in einem REST-Header

#### 3.4 Beispielimplementierung

In diesem Kapitel soll nun eine kleine Beispielimplementierung vorgestellt werden, die die oben genannte API jets3t verwendet. Diese API basiert auf REST. Allerdings ist der Zugriff auf den Service soweit abstrahiert, dass man das zugrunde liegende Protokoll gar nicht bemerkt.

In der nachfolgenden Abbildung 6 sieht man, wie man sich mit dem S3-Service verbinden kann. Der Verbindungsaufbau ist soweit gekapselt, dass man lediglich ein neues Service-Objekt anlegen muss. Hierfür ist es notwendig seinen Schlüssel anzugeben. Wird dieser akzeptiert und hat man eine Internetanbindung, hat man den Verbindungsaufbau innerhalb kürzester Zeit realisiert.

Ebenso einfach ist es sich zum Beispiel alle Buckets anzeigen zu lassen, die man verwaltet. Wie in Abbildung 7 zu sehen ist kann man über das zuvor erstellte Service-Objekt direkt eine entsprechende Methode aufrufen.

Die Funktionsweise, eine Liste der vorhandenen Objekte innerhalb eines Buckets zu erhalten ist die gleiche, nur, dass man noch einen entsprechenden Bucketnamen angeben muss. Dies ist in Abbildung 8 dargestellt.

Das Handling ist denkbar einfach. Möchte man noch mehr über die Bibliothek von jets3t erfahren, so kann man das in einer entsprechenden Dokumentation nachlesen [11].

Abbildung 6. Verbindung zum Service aufbauen

```
public String listAllBuckets() throws S3ServiceException {
    // get All Buckets
    S3Bucket[] myBuckets = s3Service.listAllBuckets();
    String myBucketString = "";

    for (S3Bucket s3Bucket : myBuckets) {
        myBucketString += "--> Name: " + s3Bucket.getName() + "\n";
        myBucketString += "--> CreationDate: " + s3Bucket.getCreationDate() + "\n";
        myBucketString += "--> Owner: " + s3Bucket.getOwner() + "\n";
    }
    return myBucketString;
}
```

Abbildung 7. Alle Buckets auflisten

Abbildung 8. Alle Objekte innerhalb eines Bucktes auflisten

#### 3.5 Kostenübersicht

Ein wichtiges Entscheidungskriterium für oder gegen den Einsatz von Amazon S3 ist der Kostenpunkt. Die Kosten für die Nutzung des Amazon S3 teilen sich in zwei Punkte auf, die Kosten für die Speicherung von Daten und die Kosten für den Transfer von Daten. Im ersten Fall unterscheiden sich die Preise außerdem, je nachdem, ob man seine Daten in Europa oder in Amerika ablegen möchte. Anhand der Zahlen in der nachfolgenden Tabelle 1 sieht man, dass Kosten erst für große Datenmengen anfangen. Die Kosten sind so gestaffelt, dass sie je größer die Datenmengen sind verhältnismäßig günstiger werden.

Tabelle 1. Kosten für den Datenspeicher [6]

GB pro Monat	USA	Europa
ersten 50 TB	\$0.150	\$0.180
nächste 50 TB	\$0.140	\$0.170
nächste 400 TB	\$0.130	\$0.160
über 500 TB	\$0.120	\$0.150

Die Kosten, die wirklich etwas ausmachen sind in den Datentransferkosten versteckt, denn je mehr Daten transferiert werden, desto teurer wird die Angelegenheit. Aber auch hier bietet Amazon eine Möglichkeit um diese Kosten zu senken. Im nächsten Abschnitt wird das Prinzip von Bittorrent erläutert und wie man dieses Protokoll in Zusammenhang mit dem S3 nutzen kann [3].

Tabelle 2. Kosten für den Datentransfer [6]

GB pro Monat	USA / Europa
ersten 10 TB	\$0.170
nächste 40 TB	\$0.130
nächste 100 TB	\$0.110
über 150 TB	\$0.100

#### 3.6 Bittorrent

Bei Bittorrent handelt es sich um eine offenes Peer-to-Peer-Protokoll zum Anbieten von Dateien. Amazon bietet die Nutzung von Bittorrent an, um öffentliche Dateien aus dem S3 anzubieten. Das hat einen entscheidenden Vorteil, nämlich den, das Kosten gespart werden können. Das funktioniert dadurch, dass jeder Nutzer nur einen Teil der Daten runter lädt und dann gleichzeitig an andere verteilt. So fungieren einzelne PCs als Server und die ganze Netzwerklast liegt nicht mehr bei Amazon. Die Schlussfolgerung daraus ist trivial: Je weniger Datentransfer auf den Servern von Amazon statt findet, desto geringer sind die Kosten für den Endnutzer.

## 4 Zusammenfassung

Am Ende dieser Ausarbeitung kann nun die Frage beantwortet werden, was Amazon S3 ist, nämlich ein flexibler Service, der es ermöglicht Daten relativ schnell an einen unbestimmten Ort zu speichern mit dem Vorteil, dass dieser Ort als Verteilungsplattform für diese Daten dienen kann. Amazon S3 zu nutzen ist nicht schwer, da viele Tools und Bibliotheken angeboten werden, um schnell auf den Service zuzugreifen. Der Einsatz ist besonders sinnvoll, wenn Systeme untereinander mit gemeinsamen Daten kommunizieren müssen. Hier kann der Service besonders bei Peaks, bei denen besonders hohe Datenmengen vorkommen eingesetzt werden. Setzt man den Service ein, um Daten für andere zur Verfügung zu stellen, sollte man die Kosten im Auge behalten, diese könnten sehr schnell sehr hoch werden.

Zu empfehlen ist der Einsatz vom S3, wenn es sich nicht um allzu geheime Daten handelt. Amazon bietet zwar einige Sicherheitsaspekte, aber letztendlich weiß der Anwender nicht, wo seine Daten liegen und was mit Ihnen geschieht. Sollten die Daten verloren gehen, entsteht vielleicht ein Schaden, der unbezahlbar ist.

### Literatur

 Reese, George. Cloud Application Architectures. 1. Auflage. O'Reilly Media. April 2009 Shankland, Stephen, KadenWillms, Jan. Gartner: Cloud Computing wird wichtigster IT-Trend 2010.

http://www.zdnet.de/news/wirtschaft\_unternehmen\_business\_gartner\_cloud\_computing\_wird\_wichtigster\_it\_trend\_2010\_story-39001020-41516155-1.htm

2009 October 21

- 3. Rubner, Stefan. Eimerweise Online-Speicher. c't 2008, Heft 23, Seite 186-187
- 4. © 2009, Amazon Web Services Amazon Elastic Block Store (EBS)

http://aws.amazon.com/ebs/

5. ©2009, Amazon Web Services What is AWS?

http://aws.amazon.com/what-is-aws/

6. © 2009, Amazon Web Services Amazon Simple Storage Service (Amazon S3)

http://aws.amazon.com/s3

Armbrust, M., A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia Above the Clouds: A Berkeley View of Cloud Computing. Publizierungsdatum: 2009.

http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html

- 8. das ELKO das Elektronikkomendium, Weblink am 30.10.2009 das ELKO
  - http://www.elektronik-kompendium.de/sites/net/0906071.htm
- 9. ©2008, Amazon Web Services LLC or its affiliates. All rights reserved. Amazon Web Services: Overview of Security Processes
  - http://developer.amazonwebservices.com/connect/entry!default.jspa?categoryID=152&externalID=1697
- 10. Amazon S3 Firefox Organizer , Version 0.4.8 September 11, 2009 –113kb S3Fox Organizer

https://addons.mozilla.org/en-US/firefox/addon/3247

11. © 1995 - 2007 Collab Net. Collab Net is a registered trademark of Collab Net, Inc.  ${\it JetS3t}$  -  ${\it Development\ Home}$ 

https://jets3t.dev.java.net/