

Inhaltsverzeichnis

1	Einleitung	1
1.1	Projektbeschreibung	5
1.2	Inspiration	6
1.3	Anforderungen	8
1.3.1	Datensicherheit	8
1.3.2	Funktionalitäten	10
1.3.3	Architektur	10
1.3.4	Weiterführende Anforderungen	11
1.4	Kapitelübersicht	11
	Literaturverzeichnis	13

Abbildungsverzeichnis

1.1	Anzahl der Dropbox-Nutzer weltweit zwischen Januar 2010 und Mai 2014 (in Millionen) [Dropbox 2014]	2
1.2	Hauptbedenken der Nutzer von Cloud-Diensten in Österreich im Jahr 2012 [Accenture 2012]	2
1.3	Zustimmung zu der Aussage: “Der NSA-Skandal hat das Vertrauen in Cloud-Dienste beschädigt.” [eco 2014]	3

1 Einleitung

Seit dem Aufkommen von Cloud-Diensten, befinden sich immer mehr AnwenderInnen in einem Konflikt zwischen Datensicherheit und Datenschutz. Cloud-Dienste ermöglichen es, Daten sicher zu speichern und mit seinen Mitmenschen zu teilen. Jedoch gibt es große Bedenken der BenutzerInnen im Bezug auf den Datenschutz, wenn sie Ihre Daten aus der Hand geben. Diese Problematik veranschaulichen auch verschiedene Studien. Sie beweisen, dass es immer mehr NutzerInnen in die Cloud zieht (siehe Abbildung 1.1), dabei aber die Bedenken gegen genau diese Anwendungen, zunehmen (siehe Abbildung 1.2).

Die Statistik aus der Abbildung 1.1 zeigt, wie die Nutzerzahlen des kommerziellen Cloud-Dienstes Dropbox¹, in den Jahren 2010 bis 2014 von anfänglich 4 Millionen auf 300 Millionen im Jahre 2014 angestiegen sind.

Im Gegensatz dazu wurde im Jahre 2012 in Österreich erhoben, dass nur etwa 17% der AnwenderInnen diese Dienste ohne Bedenken verwendet. Das meistgenannte Bedenken dieser Studie ist: **Fremdzugriff auf die Daten ohne Information.**

Dieses Bedenken ist seit den Abhörskandalen, durch verschiedenste Geheimdienste wie zum Beispiel die NSA, noch verstärkt worden. Dies zeigt eine Umfrage aus dem Jahre 2014, die in Deutschland durchgeführt wurde (Abbildung 1.3). Dabei gaben 71% an, dass das Vertrauen zu Cloud-Diensten durch diese Skandale beschädigt worden ist.

¹<https://www.dropbox.com/>

1 Einleitung

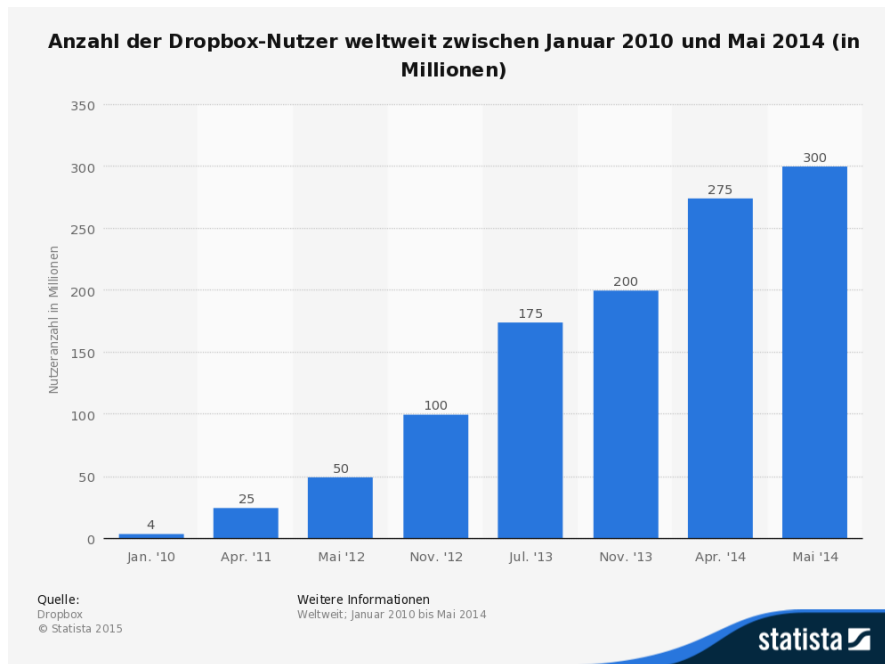


Abbildung 1.1: Anzahl der Dropbox-Nutzer weltweit zwischen Januar 2010 und Mai 2014 (in Millionen) [Dropbox 2014]

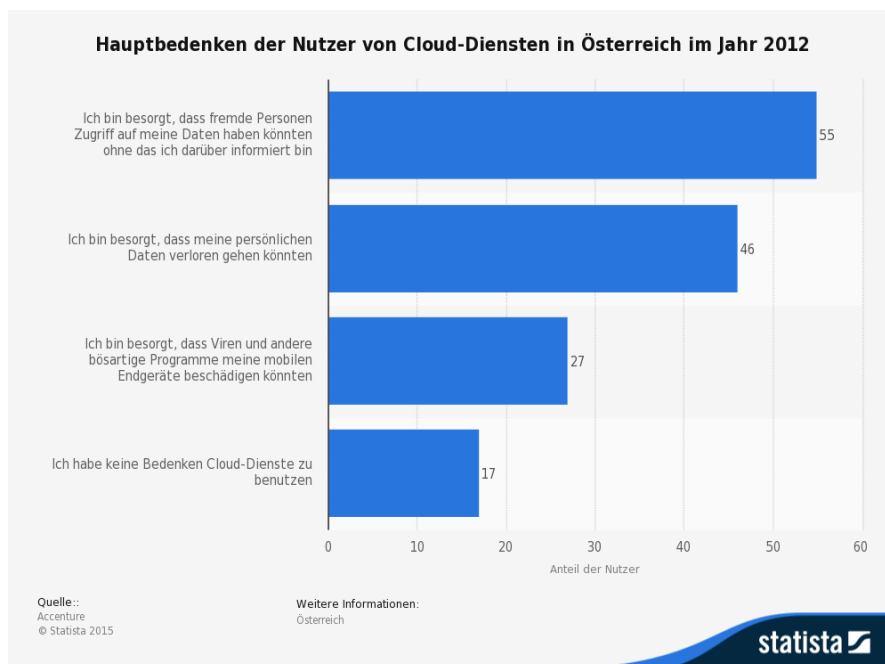


Abbildung 1.2: Hauptbedenken der Nutzer von Cloud-Diensten in Österreich im Jahr 2012 [Accenture 2012]

1 Einleitung

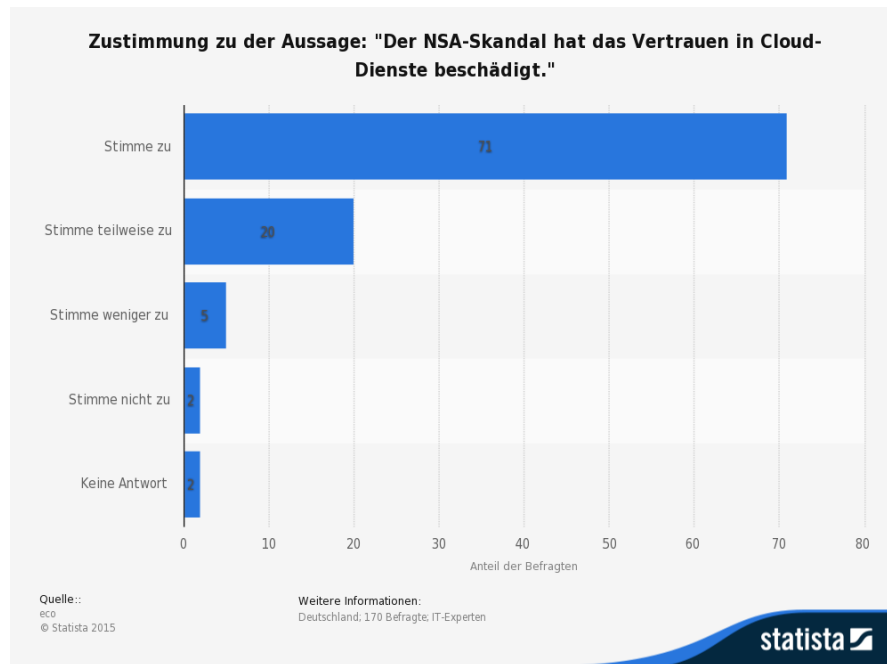


Abbildung 1.3: Zustimmung zu der Aussage: “Der NSA-Skandal hat das Vertrauen in Cloud-Dienste beschädigt.” [eco 2014]

Diese Statistiken zeigen, dass immer mehr Menschen das Bedürfnis verspüren, die Kontrolle über ihre Daten zu behalten, aber trotzdem die Vorzüge solcher Dienste nutzen wollen. Aufgrund dessen erregen Projekte wie Diaspora², ownCloud³ und ähnliche Softwarelösungen immer mehr Aufmerksamkeit.

Diaspora Diaspora ist ein dezentrales soziales Netzwerk. Die BenutzerInnen von diesem sozialen Netzwerk sind durch die verteilte Infrastruktur nicht von einem Dienstleister abhängig. Es ermöglicht, seinen Freunden bzw. der Familie, eine private “social-media” Plattform anzubieten und diese nach seinen Wünschen zu gestalten. Das Interessante daran sind die sogenannten Pods (dezentrale Knoten), die sich beliebig untereinander vernetzen lassen. Dies ermöglicht es auch Benutzern, die nicht auf demselben Server registriert sind, miteinander zu kommunizieren. Pods

²<https://diasporafoundation.org/>

³<https://owncloud.org/>

1 Einleitung

können von jedem installiert und betrieben werden; dabei kann der Betreiber bestimmen, wer in sein Netzwerk eintreten darf und welche Server mit seinem Kontakt aufnehmen dürfen. Die verbundenen Pods tauschen die Daten ohne einen zentralen Knoten aus. Dies garantiert die volle Kontrolle über seine Daten im Netzwerk [Diasporafoundation 2015]. Entwickelt wurde dieses Projekt in der Programmiersprache Ruby.

ownCloud Das Projekt ownCloud ist eine Software, die es ermöglicht, Dateien in einer privaten Cloud zu verwalten. Mittels Endgeräte-Clients können die Dateien synchronisiert und über die Plattform auch geteilt werden. Insgesamt bietet die Software einen ähnlichen Funktionsumfang gängiger kommerzieller Lösungen an [ownCloud 2015]. Zusätzlich bietet es eine Kollaborationsplattform, mit der zum Beispiel Dokumente über einen online Editor, von mehreren Benutzern gleichzeitig, bearbeitet werden können. Diese Technologie basiert auf der JavaScript Library WebODF⁴. Implementiert ist dieses Projekt hauptsächlich in den Programmiersprachen PHP und JavaScript.

Beide Software-Pakete ermöglichen es den NutzerInnen, Ihre Daten in einer vertrauenswürdigen Umgebung zu verwalten. Diese Umgebung wird nur ungern verlassen, um seine Daten anderen zur Verfügung zu stellen. Aufgrund dieses Umstandes, ist es für Anwendungen oft sehr schwer, sich für die breite Masse zu etablieren. In dieser Arbeit wird speziell auf die Anforderungen von Anwendungen eingegangen, die es ermöglichen soll, Dateien zu verwalten, zu teilen und in einem definierbaren Netzwerk zu verteilen. Speziell wird der Fall betrachtet, wenn zwei BenutzerInnen die auf verschiedenen Servern registriert sind, Dateien zusammen verwenden wollen. Dabei sollen die Vorgänge, die nötig sind, um die Dateien zwischen den Servern zu übertragen, transparent für die NutzerInnen gehandhabt werden.

⁴<http://webodf.org/>

1.1 Projektbeschreibung

SymCloud ist eine private Cloud-Software, die es ermöglicht, über dezentrale Knoten (ähnlich wie Diaspora) Dateien über die Grenzen des eigenen Servers hinweg zu teilen. Verbundene Knoten tauschen über sichere Kanäle Daten aus, die anschließend über einen Client mit dem Endgerät synchronisiert werden können. Dabei ist es für den Benutzer irrelevant, woher die Daten stammen.

Wobei es bei der Arbeit weniger um die Plattform, als um die Konzepte geht, die es ermöglichen eine solche Plattform umzusetzen. Dabei wird im speziellen die Datenhaltung für solche Systeme betrachtet. Um diese Konzepte, so unabhängig wie möglich von der Plattform zu gestalten, wird die Implementierung dieser Konzepte in einer eigenständigen Library entwickelt. Dieser Umstand ermöglicht eine Weiterverwendung in anderen Plattformen und Anwendungen, die ihren BenutzerInnen ermöglichen wollen, Dateien zu erstellen, verwalten, bearbeiten oder teilen. Damit kann das erstellte Konzept als Grundlage für eine “Spezifikation” von derartigen Prozessen weiterverwendet werden.

In der ersten Phase, in der diese Arbeit entsteht, werden grundlegende Konzepte aufgestellt. Diese beginnen mit der Festlegung eines Datenmodells und der Implementierung einer Datenbank, die diese Daten mit anderen Server teilen kann. Dieses Teilen von Daten soll voll konfigurierbar sein, dies bedeutet, dass die AdministratorInnen die Freiheit haben zu entscheiden, welche Server welche Daten zur Verfügung gestellt bekommen. Dabei gibt es zwei Stufen der Konfiguration, zum einen über eine Liste von vertrauenswürdigen Servern, welche sozusagen eine “Whitelist” darstellt, mit denen die BenutzerInnen kommunizieren dürfen. Die zweite Stufe sind die Rechte auf ein einzelnes Objekt. Diese Rechte regeln zusätzlich welche Benutzer (und damit die Server auf denen die Benutzer registriert sind) das Objekt tatsächlich verwenden dürfen.

Kurz gesagt, symCloud sollte eine Kombination der beiden Applikationen ownCloud und Diaspora sein. Dabei sollte es die Dateiverwaltungsfunktionen von ownCloud und die

Architektur von Diaspora kombinieren, um eine optimale Alternative zu kommerziellen Lösungen, wie Dropbox zu bieten.

1.2 Inspiration

Als Inspirationsquelle für das Konzept von symCloud dienten neben den schon erwähnten Applikationen auch das Projekt Xanadu⁵. Dieses Projekt wurde im Jahre 1960 von Ted Nelson gegründet und wurde nie finalisiert. Er arbeitet seit der Gründung an einer Implementierung an der Software [Atwood 2009]. Ted Nelson prägte den Begriff des Hypertext mit der Veröffentlichung eines wissenschaftlichen Artikels “The Hypertext. Proceedings of the World Documentation Federation” im Jahre 1965. Darin beschrieb er Hypertext als Lösung für die Probleme, die normales Papier mit sich bringt [Nelson ; Smith ; Mallicoat 2007].

Die darin beschriebenen Probleme sind unter anderem:

Verbindungen Text besteht oft aus einer Menge von anderen Texten wie zum Beispiel Zitate oder Querverweise. Dies lässt sich mithilfe von normalem Papier nur schwer abbilden bzw. visualisieren.

Form Ein Blatt Papier ist begrenzt in seiner Größe und Form. Es zwingt den Text in bestimmte Form, welche später weder verändert noch erweitert werden kann.

Hypertext sollte nicht das Medium sondern die BenutzerInnen in den Vordergrund stellen. Durch verschiedene Mechanismen sollte Xanadu die Möglichkeit schaffen, dass BenutzerInnen Dokumente verlinken und zusammensetzen können. Jedes Dokument wäre im Netzwerk eindeutig auffindbar und versioniert (also in verschiedenen Versionen abrufbar). Damit ist Xanadu ein nie zu Ende gebrachtes Konzept einer digitalen Bibliothek [Nelson ; Smith ; Mallicoat 2007].

⁵<http://hyperland.com/TBLpage>

1 Einleitung

1981 veröffentlichte Ted Nelson in seinem Buch “Literary Machines” 17 Thesen, die die Grundsätze des Projekts Xanadu beschreiben sollten [Nelson 1981]. Einige davon wurden durch Tim Berners-Lee in der Erfindung des Internets umgesetzt, andere jedoch vernachlässigt [Atwood 2009]. Einige dieser Thesen, die vernachlässigt wurden, sind interessante Denkanstöße für ein Projekt wie symCloud.

1. Every Xanadu server can be operated independently or in a network.

Xanadu Server können für sich alleine arbeiten oder in einem Netzwerk interagieren.

2. Every user is uniquely and securely identified.

Jeder Benutzer ist eindeutig und sicher identifizierbar.

3. Every user can search, retrieve, create and store documents.

Jeder Benutzer kann Dokumente durchsuchen, herunterladen, erstellen und speichern.

4. Every document can have secure access controls.

Jedes Dokument besitzt eine Zugriffskontrollliste. In dieser Liste wird angeführt, welcher Benutzer welche Rechte auf ein Dokument besitzt.

5. Every document can be rapidly searched, stored and retrieved without user knowledge of where it is physically stored.

Jedes Dokument kann schnell durchsucht, gespeichert und heruntergeladen werden, ohne dass der Benutzer weiß, wo das Dokument physikalisch gespeichert ist.

6. Every document is automatically stored redundantly to maintain availability even in case of a disaster.

Jedes Dokument wird redundant gespeichert, um den Verlust bei unvorhergesehenen Ereignissen zu verhindern.

Diese Thesen werden in den folgenden Anforderungen an ein System wie symCloud zusammengefasst.

1.3 Anforderungen

Aufgrund der beschriebenen Projekte, die als Inspiration verwendet wurden, werden in diesem Abschnitt die Anforderungen an ein System wie symCloud beschrieben. Diese Anforderungen sind unterteilt in:

Datensicherheit In diesem Abschnitt der Anforderungen, fallen Gebiete wie Datenschutz und der Schutz vor Fremdzugriff.

Funktionalitäten Ein System wie symCloud sollte Funktionen mit sich bringen, die es erlauben, sich gegen andere Cloud-Lösungen behaupten zu können.

Architektur Aufgrund der Inspiration durch Diaspora und Xanadu ist die Anforderung an die Architektur von verteilten Aspekten geprägt.

Neben diesen funktionalen Anforderungen gelten allgemeine Anforderungen an die Software wie zum Beispiel:

Stand der Technik Die Entwicklung der Software entspricht dem Stand der Technik.

Wartbarkeit und Erweiterbarkeit Wartbarkeit bezeichnet den Grad der Effektivität mit der eine Software, durch einen Entwickler angepasst werden kann [ISO 2011]. Gleiches gilt für die Erweiterbarkeit.

1.3.1 Datensicherheit

Sinngemäß versteht man nach DIN44300, Teil 1,

- unter **Datensicherheit** die Bewahrung von Daten vor Beeinträchtigung, insbesondere durch Verlust, Zerstörung oder Verfälschung und vor Missbrauch.
- unter **Datenschutz** die Bewahrung schutzwürdiger Belange von Betroffenen oder Beeinträchtigung durch die Verarbeitung ihrer Daten, wobei es sich bei den Betroffenen um natürliche oder juristische Personen handeln kann.

1 Einleitung

Ausgedrückt in einer weniger formalen Sprache bedeutet Datenschutz den Schutz von Daten und Programmen vor unzulässiger Benutzung [Stahlknecht ; Hasenkamp 2013].

Die internationalen Kriterien für die Bewertung der Sicherheit von Systemen gehen von drei grundsätzlichen Gefahren aus [Stahlknecht ; Hasenkamp 2013]:

Verlust der Verfügbarkeit Benötigte Daten sind, durch einen Ausfall oder Zerstörung (zum Beispiel durch einen Benutzerfehler), nicht mehr verfügbar.

Verlust der Integrität Die Daten sind unabsichtlich oder bewusst verfälscht worden.

Verlust der Vertraulichkeit Unbefugte erhalten Zugriff auf Daten, die nicht für sie bestimmt sind.

Konkrete Bedrohungen, die eines dieser drei Gefahren auslösen können, sind Katastrophen, technische Defekte oder menschliche Handlungen (ob unbewusst oder bewusst spielt hierbei keine Rolle) [Stahlknecht ; Hasenkamp 2013].

Drei der im vorherigen Abschnitt genannten Thesen des Projekts Xanadu bieten Ansätze, wie diese Anforderungen umgesetzt werden können. Durch die Redundanz (These sechs) der Daten kann sowohl der Verlust der Verfügbarkeit oder Integrität in vielen Fällen verhindert werden. Wenn das System sich vergewissern will, ob die Daten valide sind, fordert es alle Kopien der Daten an und vergleicht sie. Sind alle Versionen identisch kann eine Verfälschung ausgeschlossen werden. Nicht mehr verfügbar sind Daten erst dann, wenn alle Kopien der Daten verloren gegangen sind. Die Thesen zwei und vier bieten einen Schutz vor dem Verlust der Vertraulichkeit, indem die BenutzerIn eindeutig identifiziert werden kann und ein Zugriffsberechtigungssystem die Berechtigung überprüft. Dadurch kann ausgeschlossen werden, dass sich dritte über die Schnittstellen des Systems, Zugriff auf Daten verschaffen, die sie nicht sehen dürften.

1.3.2 Funktionalitäten

Um ein System wie symCloud konkurrenzfähig zu vergleichbaren Systemen wie Dropbox oder ownCloud [ownCloud 2015] zu machen, sind drei Kernfunktionalitäten unerlässlich:

Versionierung von Dateien Die Versionierung ist ein wesentlicher Bestandteil von vielen Filehosting-Plattformen. Es ermöglicht nicht nur das Wiederherstellen von alten Dateiversionen, sondern auch das wiederherstellen von gelöschten Dateien.

Zusammenarbeit zwischen BenutzerInnen Um eine grundlegende Zusammenarbeit zwischen BenutzerInnen zu ermöglichen, ist es unerlässlich die Dateien bzw. Ordner teilen zu können.

Zugriffsberechtigungen vergeben Um die Transparenz des Systems zu steigern, sollten die BenutzerInnen entscheiden können, welche Dateien bzw. Ordner von wem und wie verwendet werden können.

Diese drei Anforderungen sind auch Bestandteil des Xanadu Projektes. Durch die Versionierung kann sichergestellt werden, dass Dokumente wenn sie einmal veröffentlicht wurden, immer denselben Inhalt besitzen und über dieselbe URL erreichbar sind. Dies ist speziell für Zitierungen wichtig. Die Zusammenarbeit zwischen den Benutzern und die Zugriffsberechtigungen sind ebenfalls zentrale Bestandteile von Xanadu.

1.3.3 Architektur

Inspiziert von der Architektur von Diaspora sollte es möglich sein, verschiedene Installationen von symCloud zu einem Netzwerk zusammenschließen zu können. Dabei liegt der Fokus auf der Datenverteilung und nicht auf einer Lastverteilung. Dadurch können Daten gezielt im Netzwerk verteilt werden. Aufgrund der Datensicherheitsanforderungen sollten die Daten nicht wahllos im Netzwerk verteilt werden, sondern Konzepte ausgearbeitet werden, um Daten aufgrund der Zugriffsberechtigungen auf das Netzwerk zu teilen.

Eine Architektur wie die von Diaspora, erfüllt die These eins von Xanadu, indem ein Server sowohl für sich alleine arbeiten als auch in einem Netzwerk mit anderen Servern interagieren kann.

1.3.4 Weiterführende Anforderungen

Wichtige, aber in dieser Arbeit nicht betrachtete Anforderungen, sind:

Effizienz und Performance Die Effizienz und die Performance eines Systems ist meist nicht der Grund für einen Erfolg, allerdings meist eines der wichtigsten Gründe bei einem Misserfolg.

Verschlüsselung Um die Datensicherheit zu gewährleisten, sollten die Daten auf dem Speichermedium und bei der Übertragung zwischen den einzelnen Stationen, verschlüsselt werden. Um den Schutz vor Fremdzugriff auch außerhalb des Systems zu gewährleisten.

Diese Anforderungen sind, wie schon erwähnt, außerhalb des Fokuses dieser Arbeit und des Konzeptes, das während dieser Arbeit entsteht. Sie sind allerdings wichtige Anforderungen an ein produktiv eingesetztes System und sollten daher zumindest eine Erwähnung in dieser Arbeit finden. Sie sind vor allem als Anregung für weiterführende Entwicklungen oder Untersuchungen gedacht.

1.4 Kapitelübersicht

Im Kapitel ?? wird ein Überblick über den aktuellen Stand der Technik gegeben. Dabei werden zuerst einige Begriffe für die weitere Arbeit definiert. Danach werden Anwendungen und Technologien durchleuchtet, die die Bereiche Cloud-Datenhaltung, verteilte Daten und verteilte Datenmodelle umfassen.

1 Einleitung

Anschließend werden in einem Evaluierungskapitel (Kapitel ??) Technologien betrachtet, die es ermöglichen, Daten in einer verteilten Architektur zu speichern. Dazu wurden die Bereiche Objekt-Speicherdienste, verteilte Dateisysteme und Datenbank gestützte Dateisysteme mit Beispielen analysiert und auf ihre Tauglichkeit als Basis für ein Speicherkonzept evaluiert.

Das Kapitel ?? befasst sich mit der Konzeption von symCloud. Dabei geht es zentral um das Datenmodell und die Datenbank, die diese Daten speichert und verteilt.

Dieses Konzept wurde in einer Prototypen Implementierung umgesetzt. Die Details der Implementierung und die verwendeten Technologien werden in Kapitel ?? beschrieben.

Abschließend (Kapitel ??) werden die Ergebnisse der Arbeit zusammengefasst und analysiert. Zusätzlich wird ein Ausblick über die Zukunft des Projektes und mögliche Erweiterungen vorgestellt.

Literaturverzeichnis

Accenture [2012]: Hauptbedenken der Nutzer von Cloud-Diensten in Österreich im Jahr 2012. . Online im Internet: <http://de.statista.com/statistik/daten/studie/297078/umfrage/bedenken-bei-der-nutzung-von-cloud-diensten-in-oesterreich/>

Atwood, Jeff [2009]: The Xanadu Dream <http://blog.codinghorror.com/the-xanadu-dream/>.

Diasporaoundation [2015]: Was ist Dezentralisierung. . Online im Internet: <https://diasporaoundation.org/about> [Zugriff am: 05.03.2015].

Dropbox, The Next Web, TechCrunch [2014]: Anzahl der Dropbox-Nutzer weltweit zwischen Januar 2010 und Mai 2014 (in Millionen). . Online im Internet: <http://de.statista.com/statistik/daten/studie/326447/umfrage/anzahl-der-weltweiten-dropbox-nutzer/>

eco [2014]: Zustimmung zu der Aussage: Der NSA-Skandal hat das Vertrauen in Cloud-Dienste beschädigt. . Online im Internet: <http://de.statista.com/statistik/daten/studie/316667/umfrage/vertrauensverlust-bei-cloud-diensten-durch-nsa-skandal-in-deutschland/>

ISO [2011]: System and software quality models . Online im Internet: http://www.iso.org/iso/catalogue_detail.htm?csnumber=35733

Nelson, T. [1981]: Literary Machines: The Report On, and Of, Project Xanadu Concerning Word Processing, Electronic Publishing, Hypertext, Thinkertoys. XOC. Online im Internet: <https://books.google.at/books?id=LCFAXwAACAAJ>

Nelson, Theodor Holm ; Smith, Robert Adamson ; Mallicoat, Marlene [2007]: Back to the Future: Hypertext the Way It Used to Be Proceedings of the Eighteenth Conference on Hypertext and Hypermedia New York, NY, USA: ACM, S. 227–228. Online im Internet: <http://doi.acm.org/10.1145/1286240.1286303>

ownCloud [2015]: ownCloud Features. . Online im Internet: <https://owncloud.org/features> [Zugriff am: 05.03.2015].

Stahlknecht, P. ; Hasenkamp, U. [2013]: Einführung in die Wirtschaftsinformatik. Springer Berlin Heidelberg. [= Springer-Lehrbuch]. Online im Internet: https://books.google.at/books?id=9B/_0BgAAQBAJ