

Cloud Data Management

Dr. Martin Grund



CyberCamp 2013



Die Evolution des Web



■ Web 1.0:

- Entstehung des World Wide Web
- 1989 (CERN) Tim Berners-Lee.
- 1991 weltweite Verbreitung
- Navigation zwischen statischen Seiten
- Keine Interaktion zwischen den Benutzern



CyberCamp 2013

Die Evolution des Webs



■ Web 2.0

- Internetzugang wird omnipräsent
- Dynamische Seiten
 - ASP (Active Server Page)
 - PHP
- Interaktion zwischen Benutzern
 - Chat
 - Forum
 - Persönliche Nachrichten



CyberCamp 2013

Die Evolution des Webs

■ Web 2.0

- Benutzer produzieren und konsumieren Informationen
- Soziale Netzwerke
 - Facebook
 - Twitter
 - LinkedIn
 - etc.



CyberCamp 2013

Die Evolution des Webs



■ Web 3.0

➤ Web der Dinge

- RFID
- Barcode
- Augmented Reality

➤ Semantic Web

- Programme interagieren ohne menschliche Aufsicht
- RDF (Ressource Description Framework)
- OWL (Ontologie)
- etc.



CyberCamp 2013

Die Evolution des Webs

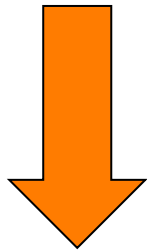


- Was für Folgen hat diese Evolution?
 - Einfacher Inhalte bereitzustellen:
 - »Jede Minute werden 60 Stunden Videomaterial auf Youtube geladen." –Youtube Press
 - Unmengen an Daten werden jeden Tag produziert
 - Benutzer geben viele mehr Informationen über sich preis
 - Es werden immer neue Wege gesucht diese Daten gewinnbringend zu verarbeiten



CyberCamp 2013

Eine Internet Minute



Technologie?
Cloud!



CyberCamp 2013





CLOUD DATA MANAGEMENT

CyberCamp 2013



UNIVERSITÉ DE FRIBOURG / DEPARTMENT OF INFORMATICS

department of informatics

Was ist die Cloud?



- Entkoppelung von Dienst und und konkrete Implementation aus Sicht der Benutzer
- Wandel von klassischer Software-Lizensierung hin zu flexiblen Preismodellen die nach Benutzung abgerechnet werden
 - Dienst werden nach Dauer, Stückzahl, Ressourcennutzung, usw. verrechnet
 - Gratis = Persönliche Daten gegen Dienstleistung

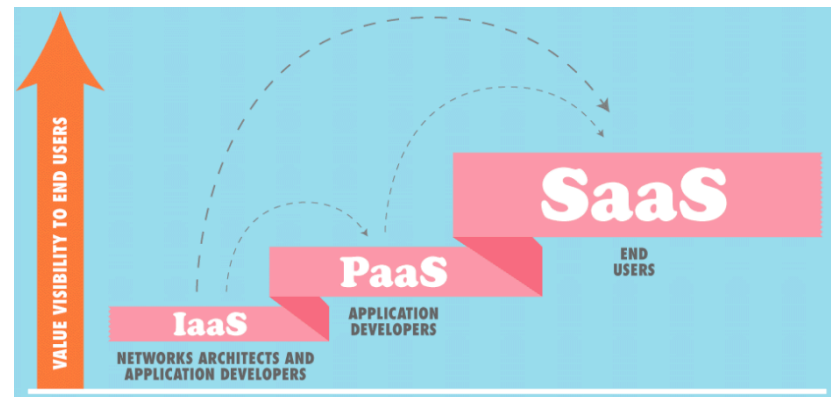


CyberCamp 2013

Was ist die Cloud

■ Verschiedene Ebene

- **Infrastructure as a Service** – Bereitstellen von Infrastrukturen
- **Platform as a Service** – Amazon EC2, bereitstellen von Rechenleistung
- **Software as a Service** – Bereitstellen von Diensten



CyberCamp 2013

Was ist die Cloud



■ Beispiele:

- Eure E-Mail Dienste(GMail, YMail).
- Eure Arbeitsdokumente. (Google Docs, Dropbox)
- Euer Quellcode (github.com, Google code).
- Euer Rechner(Amazon Cloud, Chrome PC)
- ... Auch Ihr und euere Freunde seid in der Cloud.



CyberCamp 2013

Wie findet man Dinge in der Cloud



- Suchmaschinen indexieren Milliarden von Seiten.
- Empfehlungssysteme.
- Informationsaggregatoren, z.B. RRS Feeds.



CyberCamp 2013

Plan für Heute

- Einführung in Software-as-a-Service
 - Yahoo Pipes
- Einführung in Platform-as-a-Service
 - Map-Reduce im Browser



CyberCamp 2013

Einführung in RSS Feeds



- Immer wie mehr Informationen (Blogs, Zeitungen, Foren etc.)
- Die Übersicht behalten wird zunehmend schwieriger
- Wie wäre es, wenn die Informationen zu euch kommen, anstatt dass Ihr die Informationen sucht?



CyberCamp 2013

Einführung in RSS Feeds



- Manche Seiten bieten ihre Informationen als RSS Feeds an.
- Push-Verfahren: Man braucht sich nur anzumelden um die Informationen zu erhalten.
- Verschiedene Feeds können kombiniert werden.
- RSS Feeds können mit einem RSS Reader gelesen werden.



Einführung in RSS Feeds



CyberCamp 2013

Informationen aus sozialen Netzwerken



- Durch die Verbreitung von sozialen Netzwerken werden Informationen schneller propagiert.
- Einfacher « like » Knopf reicht damit Informationen sich viral verbreiten

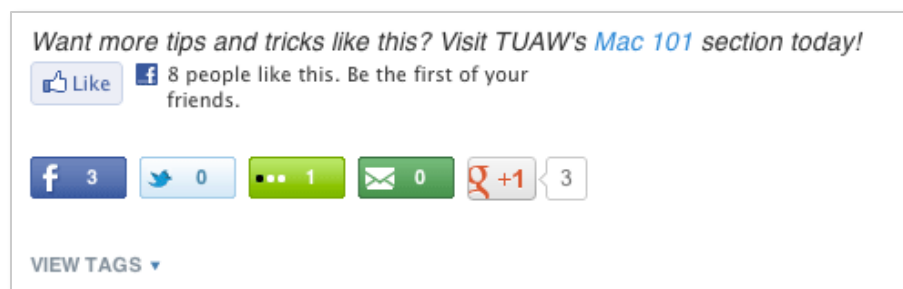


CyberCamp 2013

Informationen aus sozialen Netzwerken



- Ein « soziales » Plugin ist Programmcode, welches in HTML eingebettet werden kann.
- Webseiten werden so zunehmend in ein soziales Netzwerk eingebunden
 - Warum ?



CyberCamp 2013

Yahoo! Pipes



- Yahoo! Pipes ist ein kostenloser Dienst, um verschiedene Daten aus dem Web zu kombinieren (RSS, XML, etc.).
- Dadurch wird ein personalisierter Informationsstrom erstellt, der auch anderen Benutzern zur Verfügung gestellt werden kann.

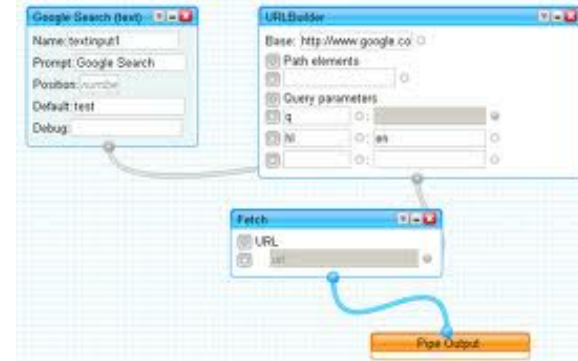


CyberCamp 2013

Yahoo! Pipes

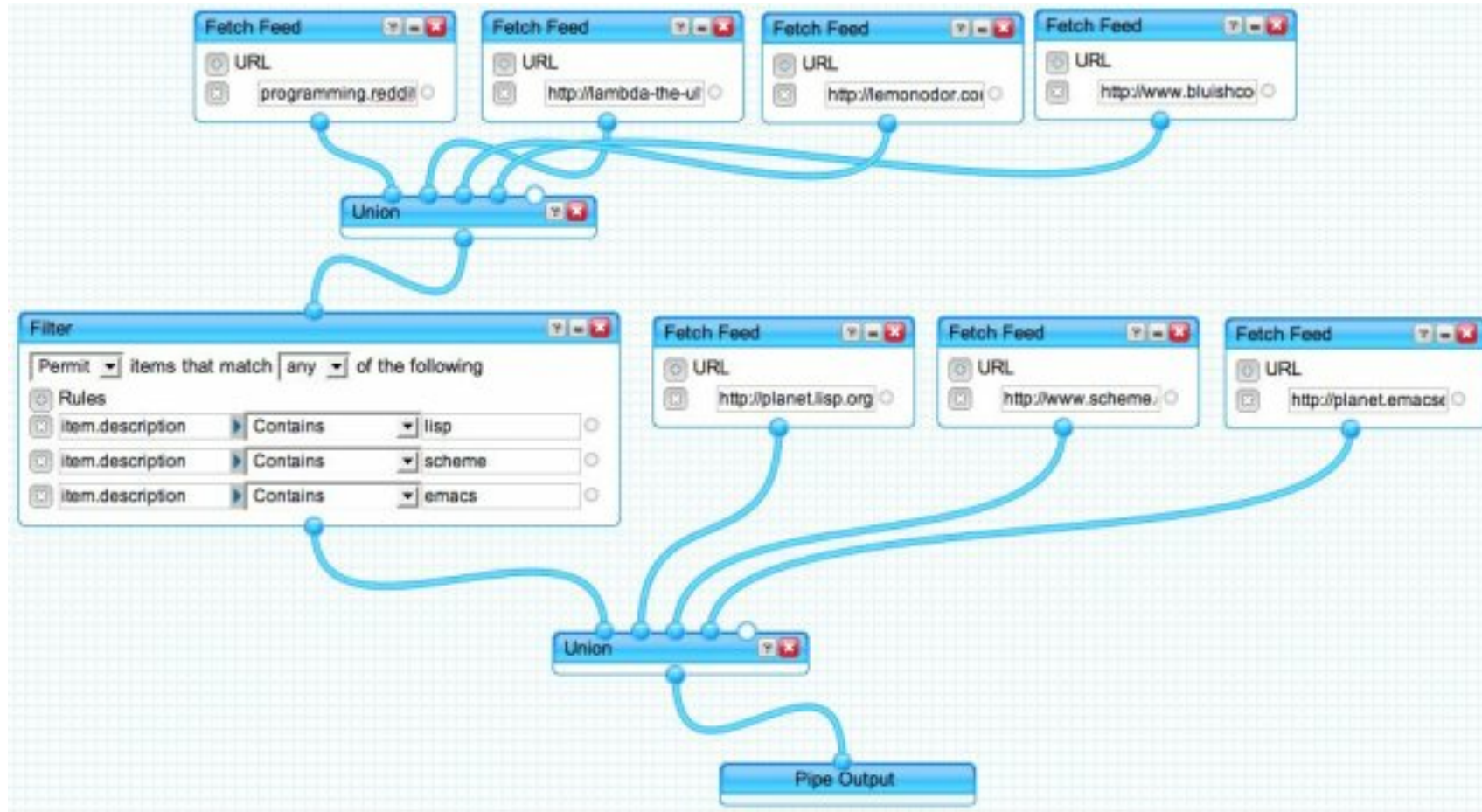


- «Pipes Editor» ist eine graphische Entwicklungsumgebung für Yahoo-Pipes um:
 - Mehrere Informationsquellen einbinden.
 - Informationen filtrieren und transformieren.
 - Informationsstrom umzuleiten.



CyberCamp 2013

Yahoo! Pipes



Massive Parallel Verarbeitung von Cloud Daten

MAP REDUCE



CyberCamp 2013

Map Reduce

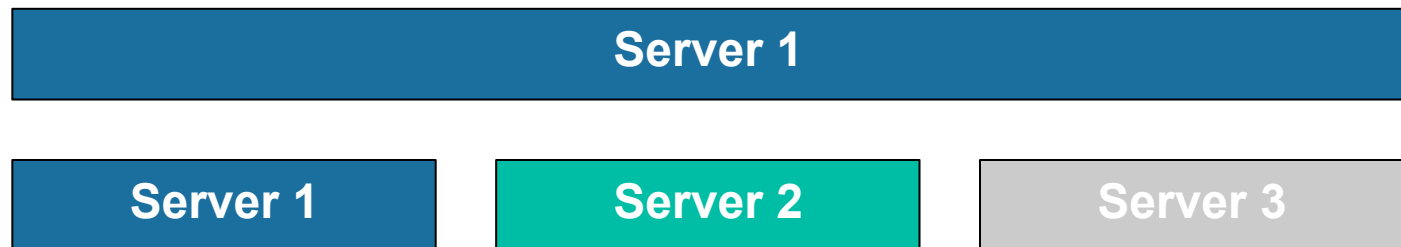
- Jeden Tag werden unvorstellbare Mengen an Daten in den Rechenzentren von Google, Facebook & Co produziert
 - Facebook hat in seiner Infrastruktur ca 180.000 Server
 - Facebooks HDF cluster hatte im August 2012 100PB Speicherplatz ~ 200.000 500GB Festplatten



CyberCamp 2013

Map Reduce

- Wie können solche unglaublichen Datenmengen verarbeitet werden?
 - 100MB/s sequentielles Lesen einer Festplatte
 - $100\text{PB}/100\text{MB/s} \Rightarrow \mathbf{32 \text{ Jahre!}}$
- Ziel: Massive parallele Verarbeitung
 - Datenparallelisierung



CyberCamp 2013

Map Reduce – Wofür?

- Stimmungsanalyse auf Twitter...
- Vorschlägen von Freunden auf Facebook...
- Analyse von wissenschaftlichen Experimenten (z.b. am CERN)...
- Analyse von Log Date zur Verhinderung von Betrug...



CyberCamp 2013

Map Reduce

*Map Reduce ist ein Programmiermodell, das es erlaubt massive **nebenläufig große Datenmenge** zu verarbeiten. Dabei wird der Prozess in zwei Schritte unterteilt. In der **MAP** Phase kann gefiltert und sortiert werden und in der **REDUCE** Phase werden Aggregationen vorgenommen.*



CyberCamp 2013

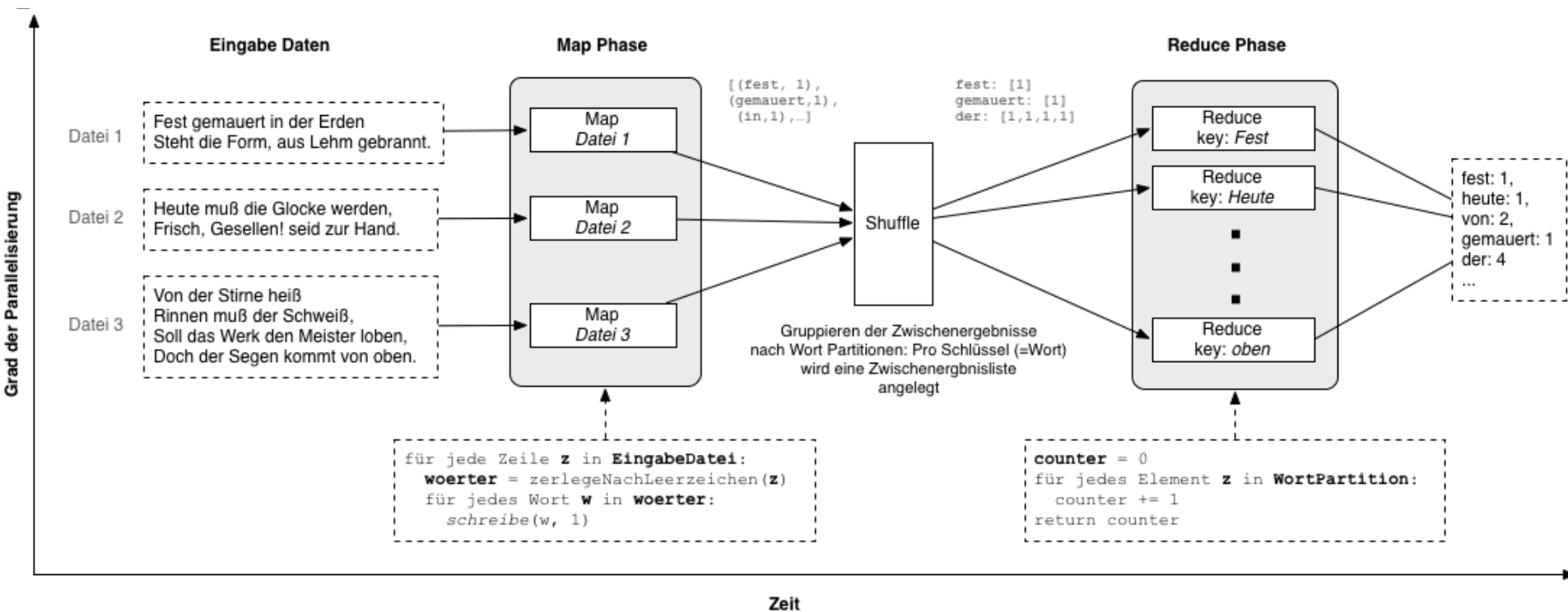
Map Reduce

- `map()` – Auf eine Liste von Eingabedaten wird eine Funktion angewendet so dass die Daten transformiert werden.
 - Aber: `size(Eingabe)` muss nicht `size(Ausgabe)` sein
 - Bsp: `map(["abc", "cdef", "a"]) { |e| e.size } = [3,4,1]`
- `reduce()` – Auf eine Liste von Eingabedaten wird eine Funktion angewendet, so dass genau ein Ausgabelement entsteht
 - Bsp: `reduce([3,4,1]) { |m,v| m + v } = 8`



CyberCamp 2013

Map Reduce Beispiel



CyberCamp 2013

Übungen

- Yahoo Pipes
 - Erstellen einer Pipe, die die Webseite von lematin.ch nach etwas durchsucht und das Ergebnis als RSS Feed ausgibt!
- Map Reduce
 - Twitter Daten analysieren!



CyberCamp 2013