CAS Software AG

Richtig informiert. Jederzeit und überall.

Verarbeitung und Visualisierung von BigData im Education-Umfeld

„Use-Cases und deren Umsetzungen bei der CAS Software AG"

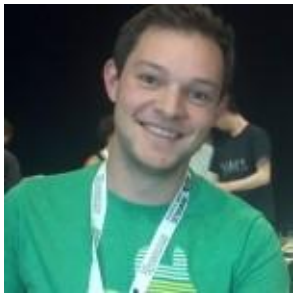
Christian Krakowski und Jonas Freiknecht, Karlsruhe 20.02.2014

Über uns

Christian Krakowski

- Dipl.-Inform.
- Bei CAS seit 2000
- Leiter Produktmanage-ment CAS Education
- christian.krakowski@cas.de

Jonas Freiknecht



- M. Sc. Informatik
- Seit 2009 IT-Spezialist bei IBM
  - Verteilte Systeme und mobile Architekturen
  - Seit 2014 Big-Data als Schwerpunkt
- jonas.freiknecht@de.ibm.com

Agenda

- CAS Education - Einführung
- (Was ist Big-Data?)
- Big-Data im Hochschulumfeld
- Anforderungen an Campus-Management-Systeme
- Auswahl eines Visualisierungs-Frameworks
- Wo kann Visualisierung unterstützen?
- Info-Grafiken
- Big-Data-Visualisierung
- Szenario 1-4
- Hands-On-Demo

CAS Education - Einführung

- Wir entwickeln und vermarkten Lösungen für die Verwaltung von Hochschulen und Schulen jeder Art
  - Campus-Management, Studierenden-Lebenszyklus, Schulverwaltung

- Das Geschäftsfeld ist vierzehn Jahre alt, aktuell bedienen wir 17 Hochschulen und etwas mehr als 100 Schulen im Inland

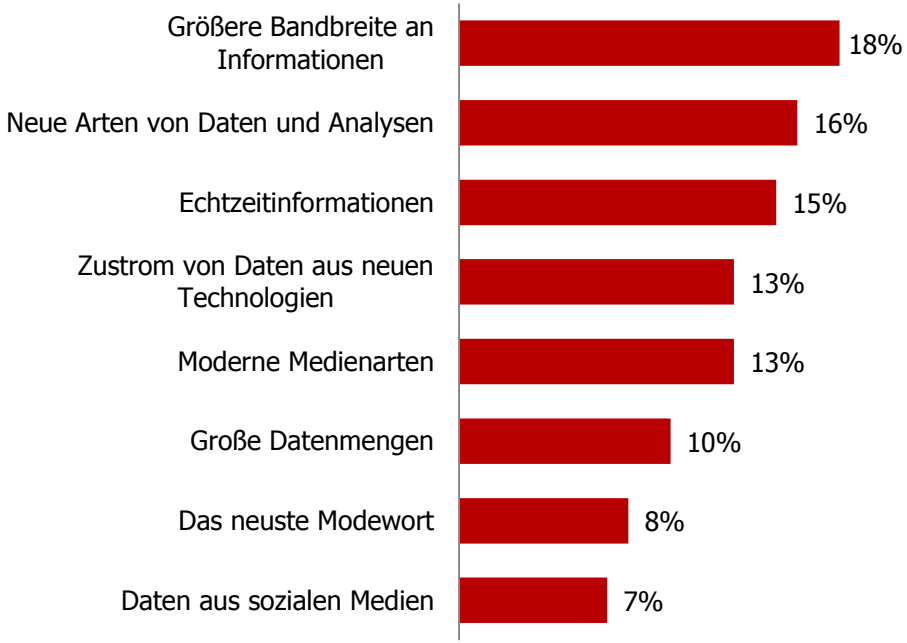
- Unsere Organisation beschäftigt ca. 30 Mitarbeiter
- Wir decken alle Prozesse im Geschäftsfeld ab, also Marketing und Vertrieb, Produktmanagement, Produktentwicklung, Projektmanagement und Kundensupport

- Eigenschaften unserer Produkte:
  - Integrierte Software
  - Standardsoftware mit erprobten Referenzprozessen
  - Flexible, anpassbare Lösungen
  - Moderne Web-Clients und mobile Clients

Once again: Was ist Big-Data?

Volume | Velocity | Variety | **Veracity**

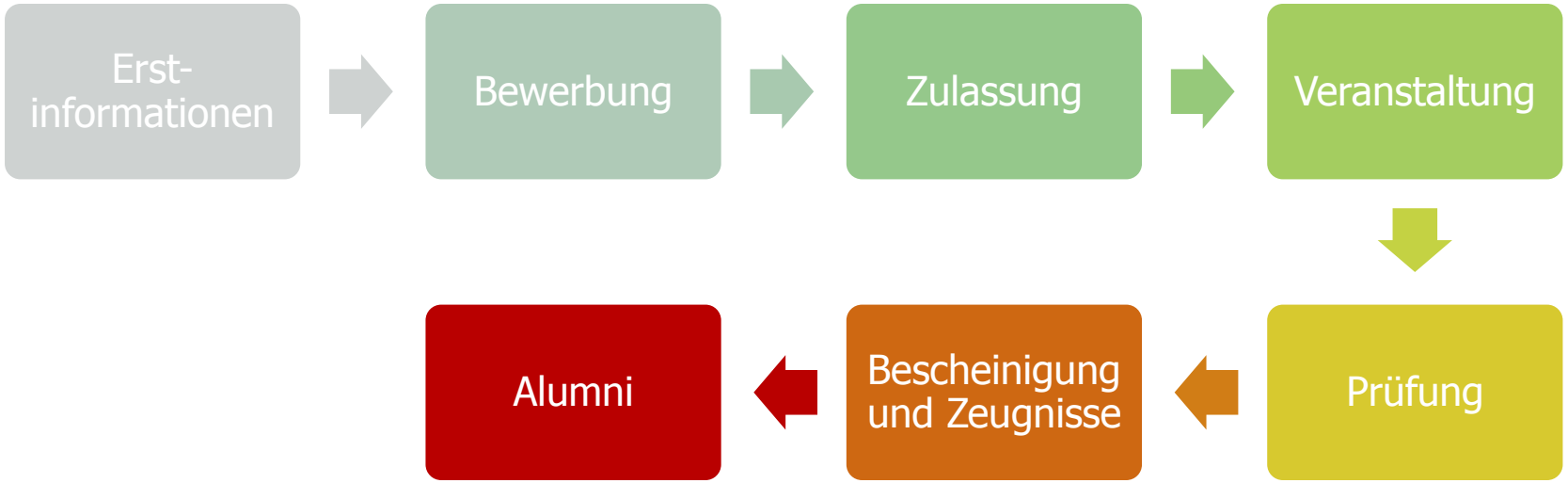
- Nicht nur Datenmenge im Fokus
- Verarbeitung ist *big*
- <u>Kein</u> Standard für neue Analysemethoden



Umfrage 2012

Big-Data im Hochschulumfeld 1/2

- Steigende Zahl von Studierenden
  - Datenwachstum im studentischen Lebenszyklus



- Immer mehr elektronische Dokumente
  - Kursdaten, Stammdaten, Präsentationen, Ausarbeitungen
- Anbindung von Learning-Management-Systemen (Moodle, Ilias, …)
- Anbindung von mobilen Geräten und sozialen Netzwerken

Big-Data im Hochschulumfeld 2/2

Szenarien

- Drop-Out-gefährdete Studenten finden
- Raumoptimierung
- Automatisierte Lehrevaluation
- Individuelle Weiterbildungsangebote
- Bewerbervorhersagen
- Marketing (soziale Netzwerke)

Anforderung an Campus-Management-Systeme (CaMS)

- Frühzeitig agieren statt reagieren zu müssen
  → Big-Data-Strategie entwickeln

- Unser Part: Bereitstellen einer Visualisierungskomponente für Web-Anwendungen (Java und ASP)

Auswahl eines Visualisierungs-Frameworks

Anforderungen

- Dynamik
- Diagrammvielfalt
- Ästhetik
- Erweiterbar
- Im Browser darstellbar
- Geringe Komplexität*

* Kein R-(Node)

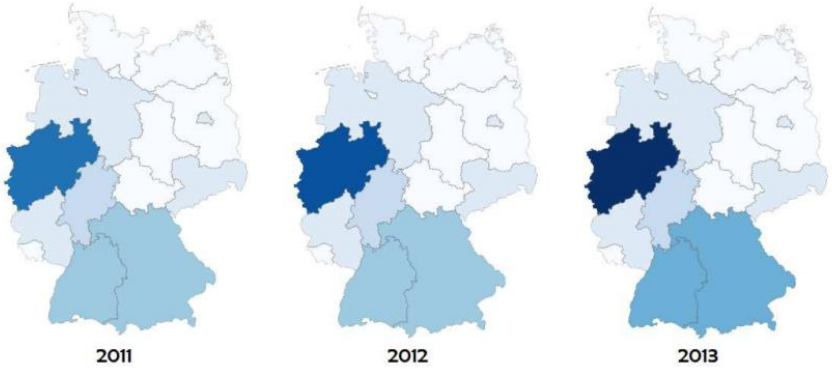
D3.js	JavaScript	SVG	~130	9	http://d3js.org/

- D3 = *Data Drive Documents*
- Entwickelt von **Mike Bostock**
- Aktuell **Version 3.3.13**
- Basiert auf HTML, JavaScript und CSS
- Vorgänger: *Protovis*
- Bekannte Nutzer: **Datameer**, The New York Times, …

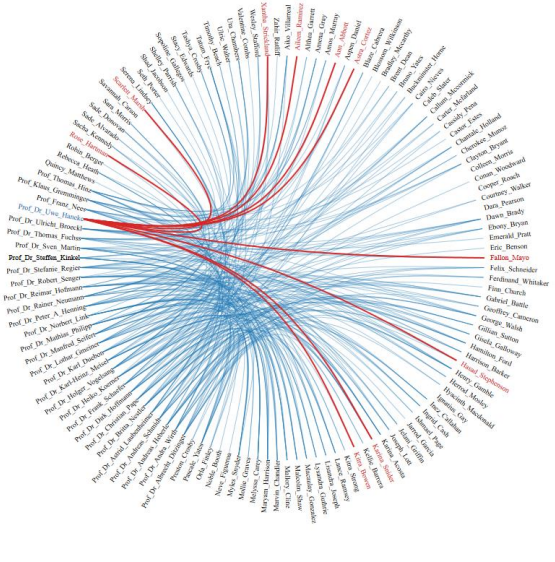
Wo kann Visualisierung im CaMS unterstützen?

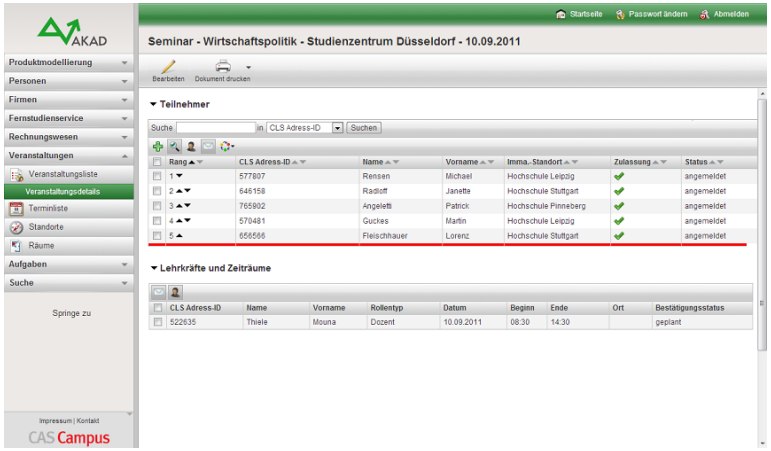
- Visualisierungen ergänzen tabellarische Darstellungen in der Verwaltung
- Im Marketing und repräsentativen Umfeld als *Eye-Catcher*

- Visualisierungen wecken Aufmerksamkeit, sind aber ungenauer



Anzahl Studierender in allen Bundesländern Deutschlands von 2011 bis 2013

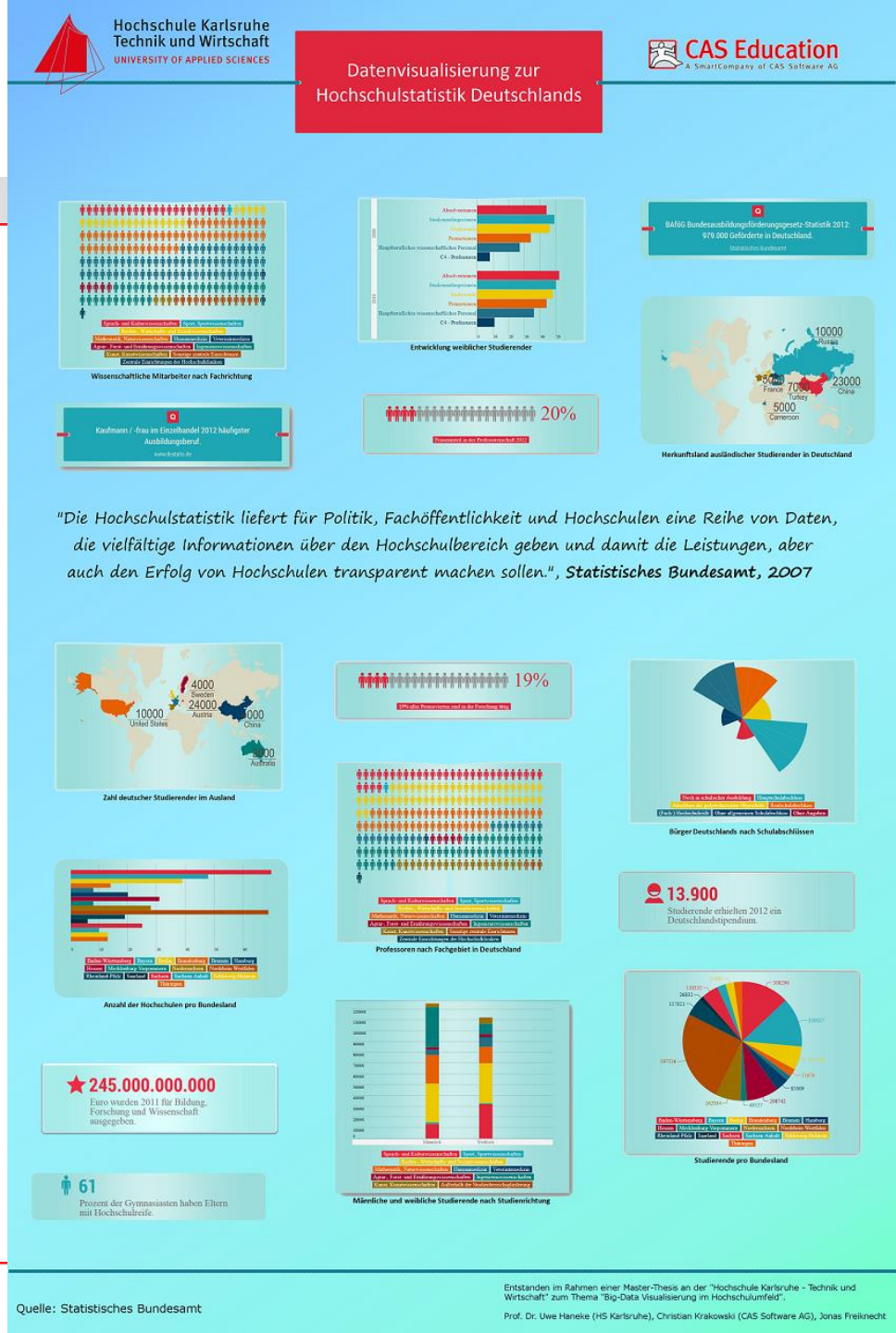




Info-Grafiken

- Zusammenstellung mehrerer Diagramme
- Alle im selben Kontext
- Zeigen häufig:
  - Klartexte
  - Prägnante Kennzahlen
  - Einfache Diagramme
- Dienen als Blickfang

http://infogr.am/



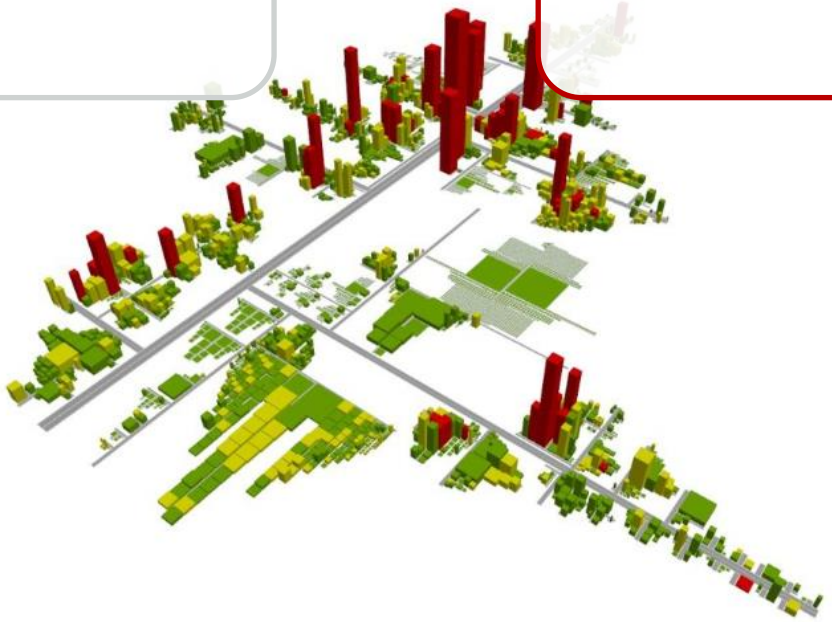
Big-Data Visualisierung

Herausforderungen

- Mehr Daten auf gleichem Raum
- Komplexe Beziehungen
- Rendern auf Client-Seite schwierig → Rechenleistung des Zielsystems unbekannt
- Verständnis komplexer Daten
- MapReduce's HashMap-Mentalität

Lösung

- Interaktivität
- Dynamik
- Kontextsensitivität
- Mehrere Dimensionen
- Vorherige Ausdünnung / Aufbereitung
- Bekannte Kontexte verwenden
- YARN



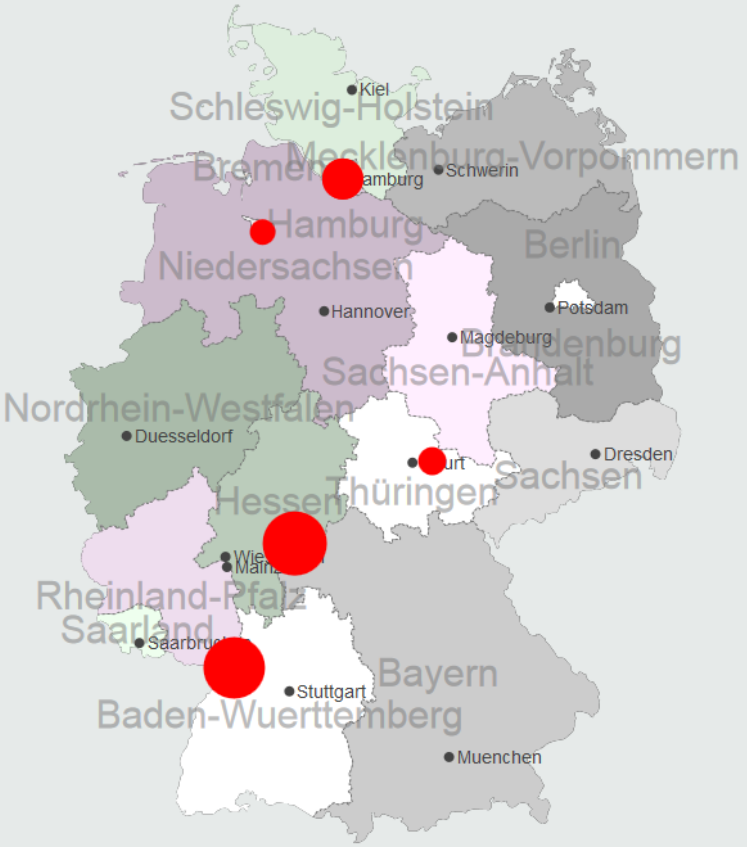
Szenario 1 – Text-Mining von Bewerberschreiben

HOME ZEIGE DATEN LADE DATEN LÖSCHE DATEN

Tabelle 1: Auswertung von Bewerberschreiben

DATEINAMEN	AUTOR	DATUM	TYP	STADT	GESCHLECHT	FEHLER
bewerbung1.docx	Jonas Freiknecht	24.10.2013 20:18:50	docx	Karlsruhe	male	20
bewerbung2.docx	Daniel Senkrecht	24.10.2013 20:18:50	docx	Bremen	male	6
bewerbung3.docx	Nina Nölle	24.10.2013 20:18:50	docx	Hamburg	unknown	12
bewerbung4.docx	Armin Rhode	24.10.2013 20:18:51	docx	Kassel	unknown	21
bewerbung5.docx	Bertold Brecht	24.10.2013 20:18:51	docx	Weimar	unknown	7
bewerbung6.docx	Carlos Caludes	24.10.2013 20:18:51	docx	Hamburg	unknown	12

Criteria: Mistakes



Szenario 2 – Komplexe M-N-Beziehungen

HOME ZEIGE DATEN

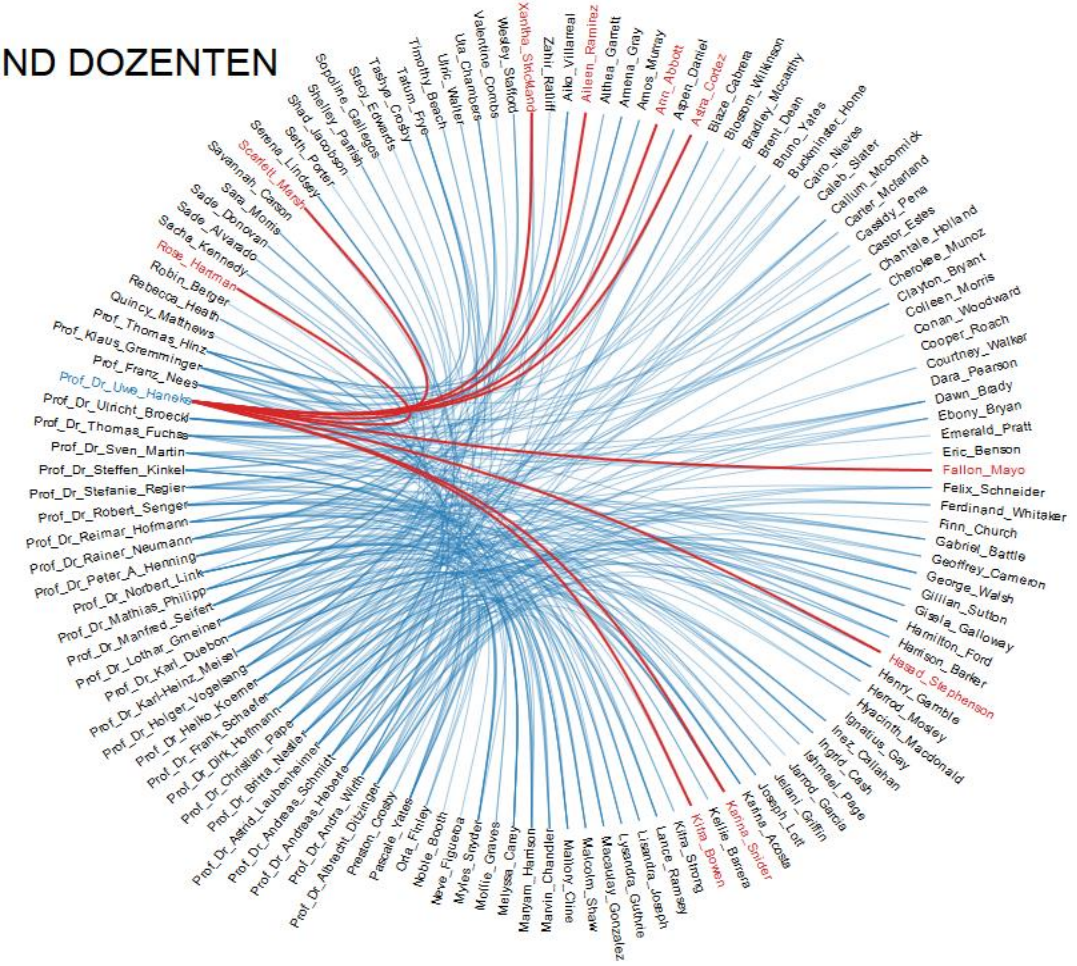
KOOPERATION VON STUDENTEN UND DOZENTEN

VISUALISIERUNG KOMPLEXER BEZIEHUNGEN

BESONDERHEITEN

- Visualisiert Zusammenarbeit von Studierenden und Dozenten (132 Datensätze).
- Größere Menge an Personen als der Masterstudiengang Informatik an der HSkA.
- Selten können Diagramme m-n-Relationen abbilden

URSPÜNGLICHE SICHT



Szenario 3 – Sentiment Analysis zur Meinungsforschung

HOME ZEIGE DATEN

TWEETS ÜBER HOCHSCHULEN

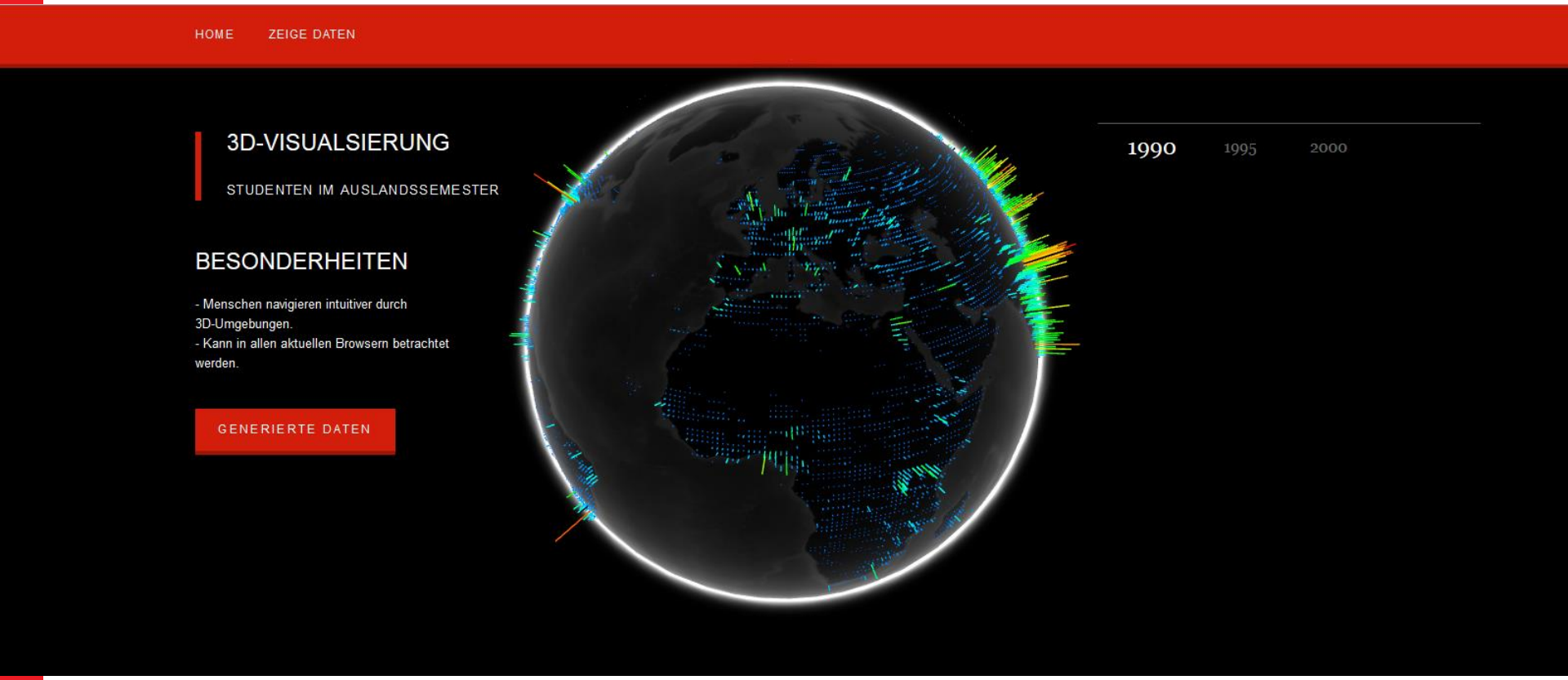
SENTIMENT ANALYSIS FÜR MEINUNGSFORSCHUNG

BESONDERHEITEN

- Extraktion von Schlagworten.

hska Visualize

Szenario 4 – 3D-Visualisierung im Browser



Hands-On-Demo

Komplexe M-N-Beziehungen

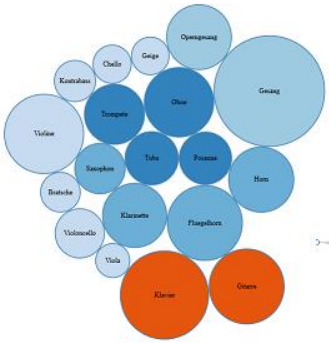
3D-Koordinaten

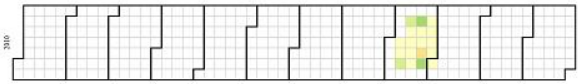
Drill-Down-Funktionalität

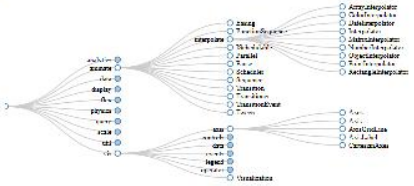
Ich bitte um Ihre Fragen

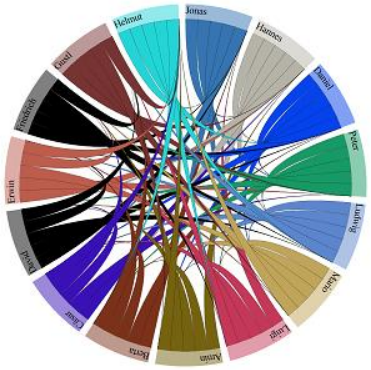
Vielen Dank für Ihre Aufmerksamkeit

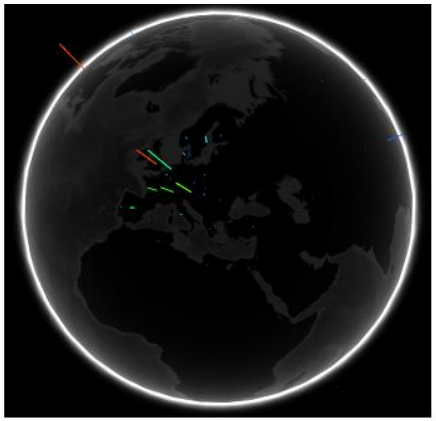
Verwendete Diagramme im Rahmen dieser Thesis

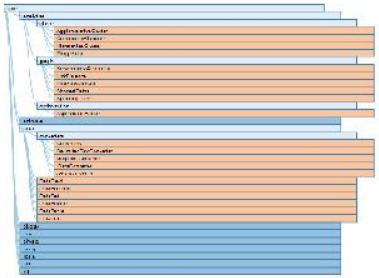


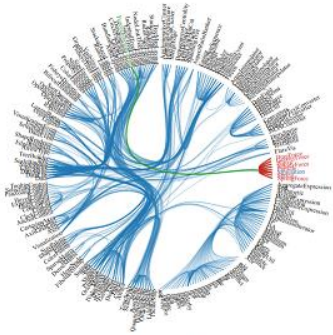


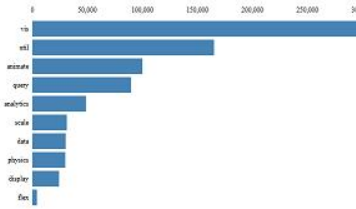


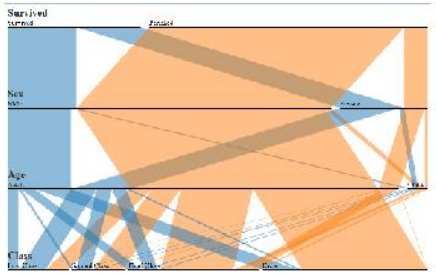


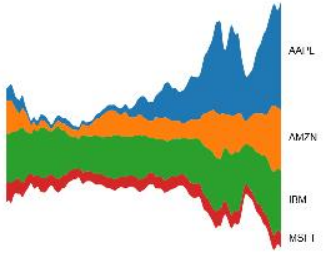


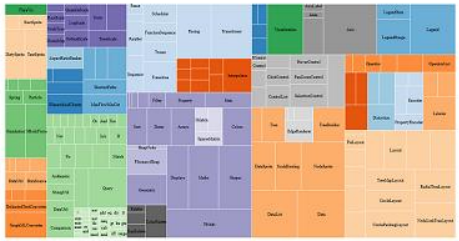




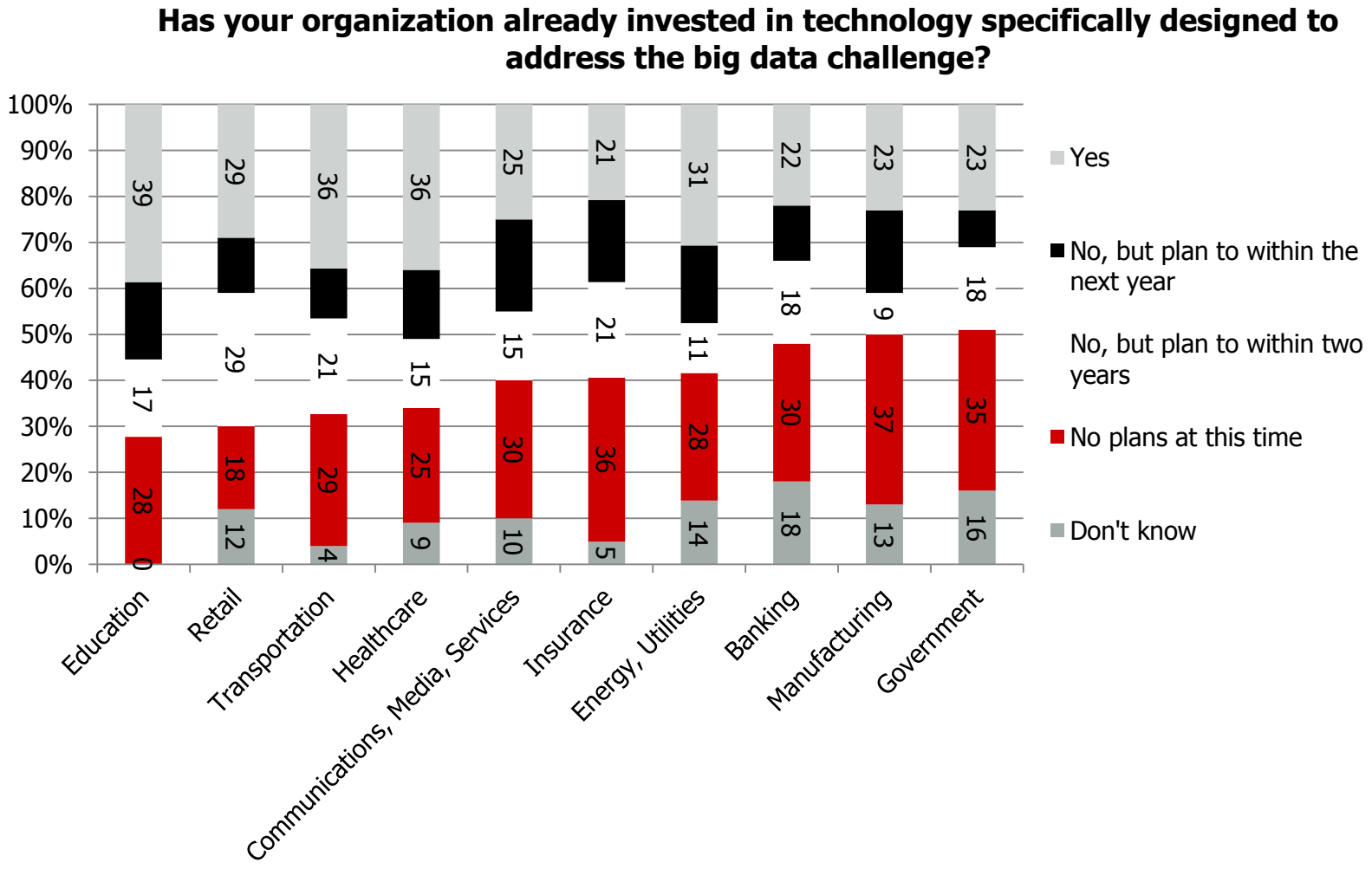




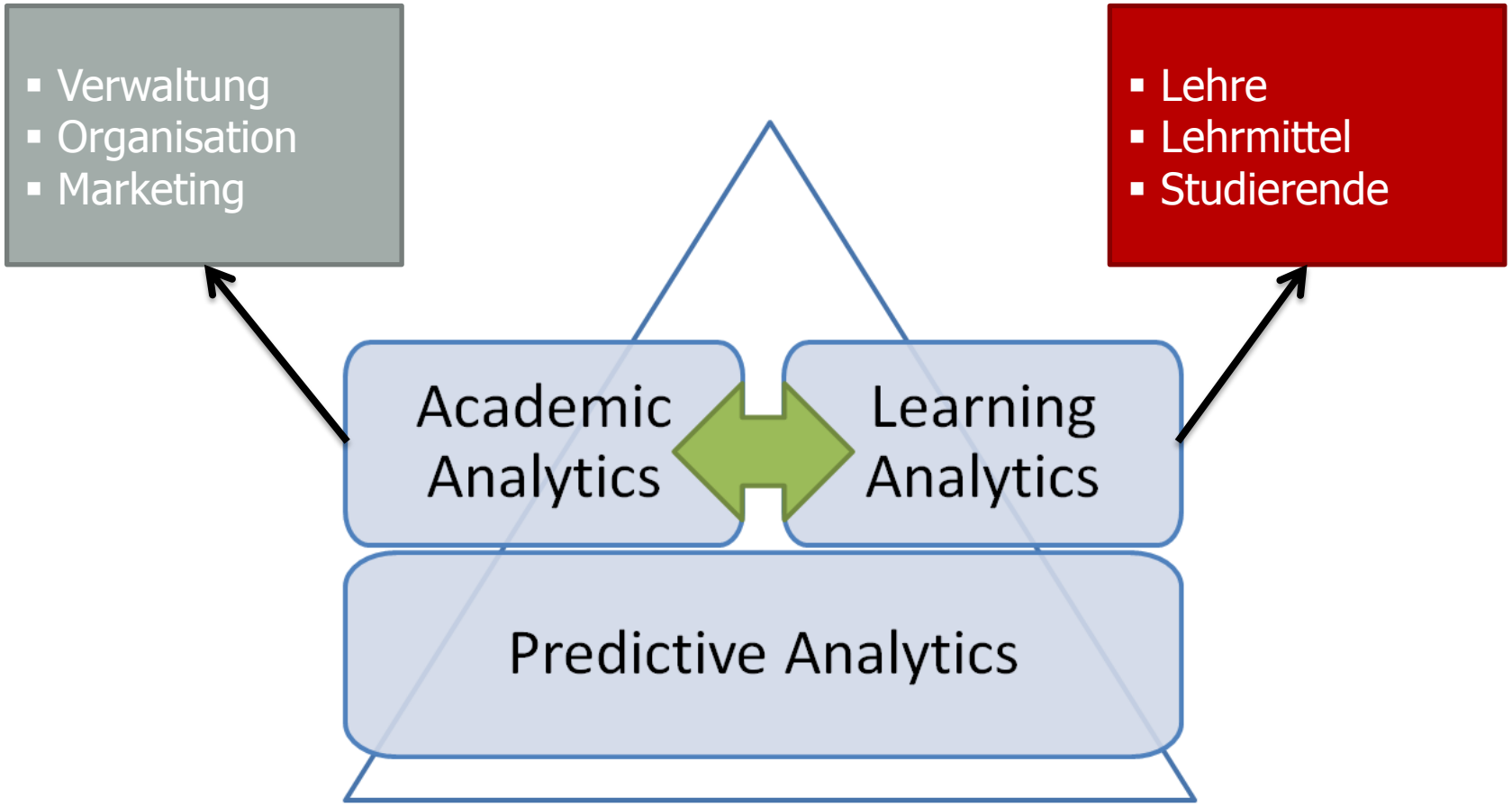




Investments in Big-Data

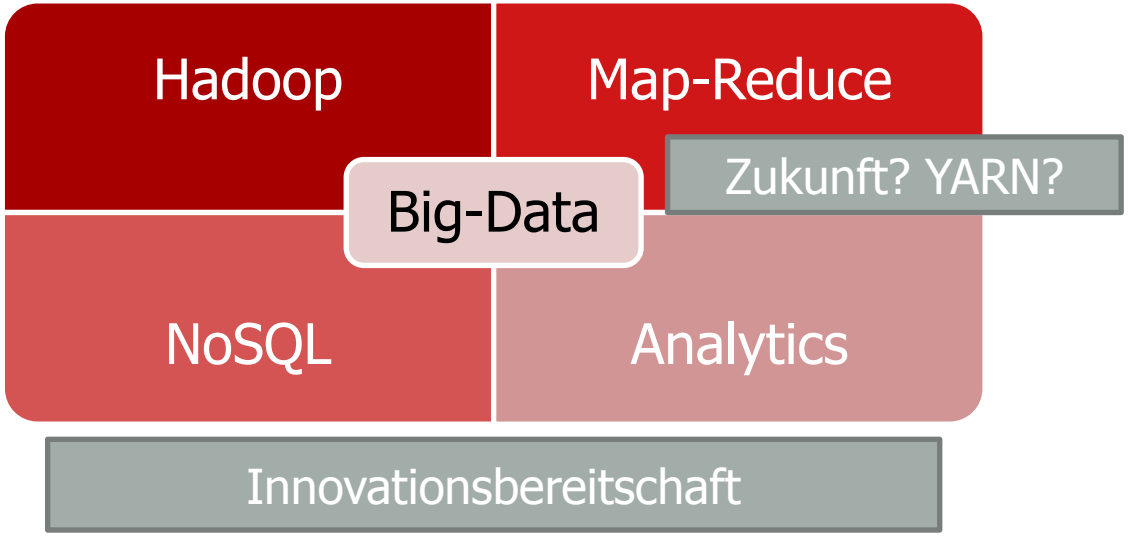


Analytische Prozesse im Hochschulwesen



Big-Data in der Praxis

- Big-Data nur, wenn dringender Bedarf besteht
  - **Herausforderungen**: Setup und Verlassen der relationalen Welt

- Erinnerung „Verarbeitung ist *big*“

- Nicht nur Hadoop oder Map-Reduce macht Big-Data aus