

Living Lab Big Data

Konzeption einer Experimentierplattform



Dr. Michael May

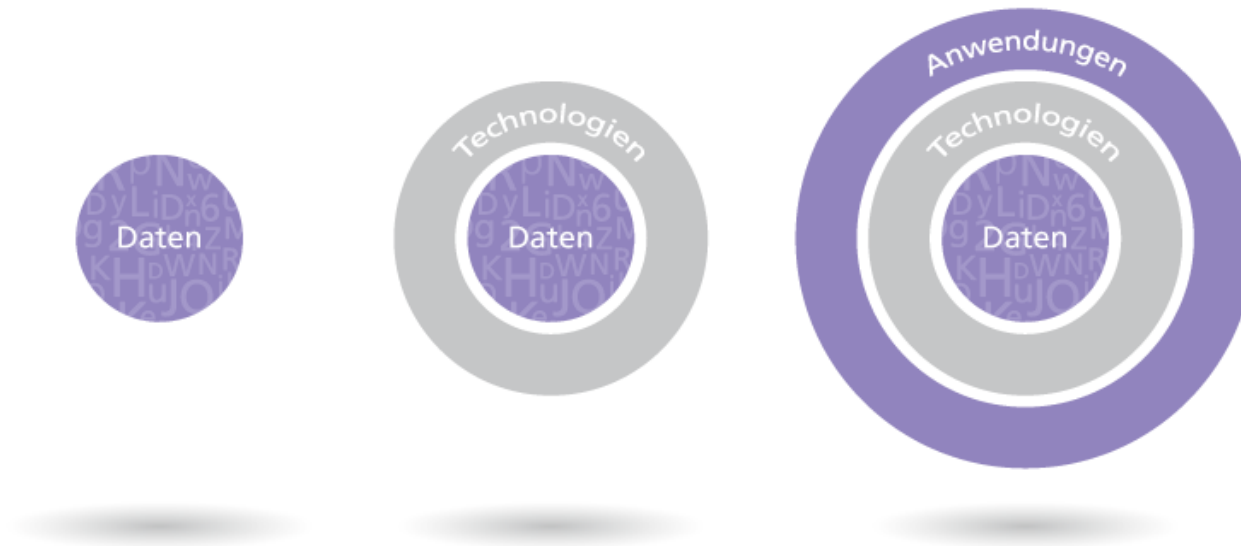
Berlin, 10.12.2012

Fraunhofer-Institut für Intelligente Analyse-
und Informationssysteme IAIS

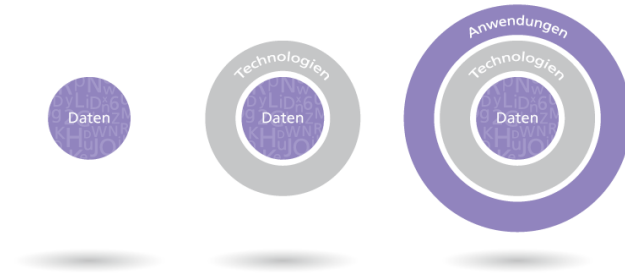
www.iais.fraunhofer.de

Agenda

- Ziele des Living Lab Big Data
- Living Lab Big Data Architektur
- Living Lab Big Data Anwendungen
- Ausblick



Ausgangslage



- Viele Anwenderunternehmen sind in der Orientierungsphase
- Vorstellungen von Big Data in Unternehmen oft diffus, sowohl zu
 - Einsatzmöglichkeiten als auch zu
 - vorhandenen Techniken
- Ergebnisse Befragung und Zukunftsworkshops:
 - Hoher Bedarf an Best Practices, Training, Schulung (95% der Befragten)
- Infrastruktur und Datenmanagement sind für sich genommen zu abstrakt, Lösungspotentiale zu bewerten

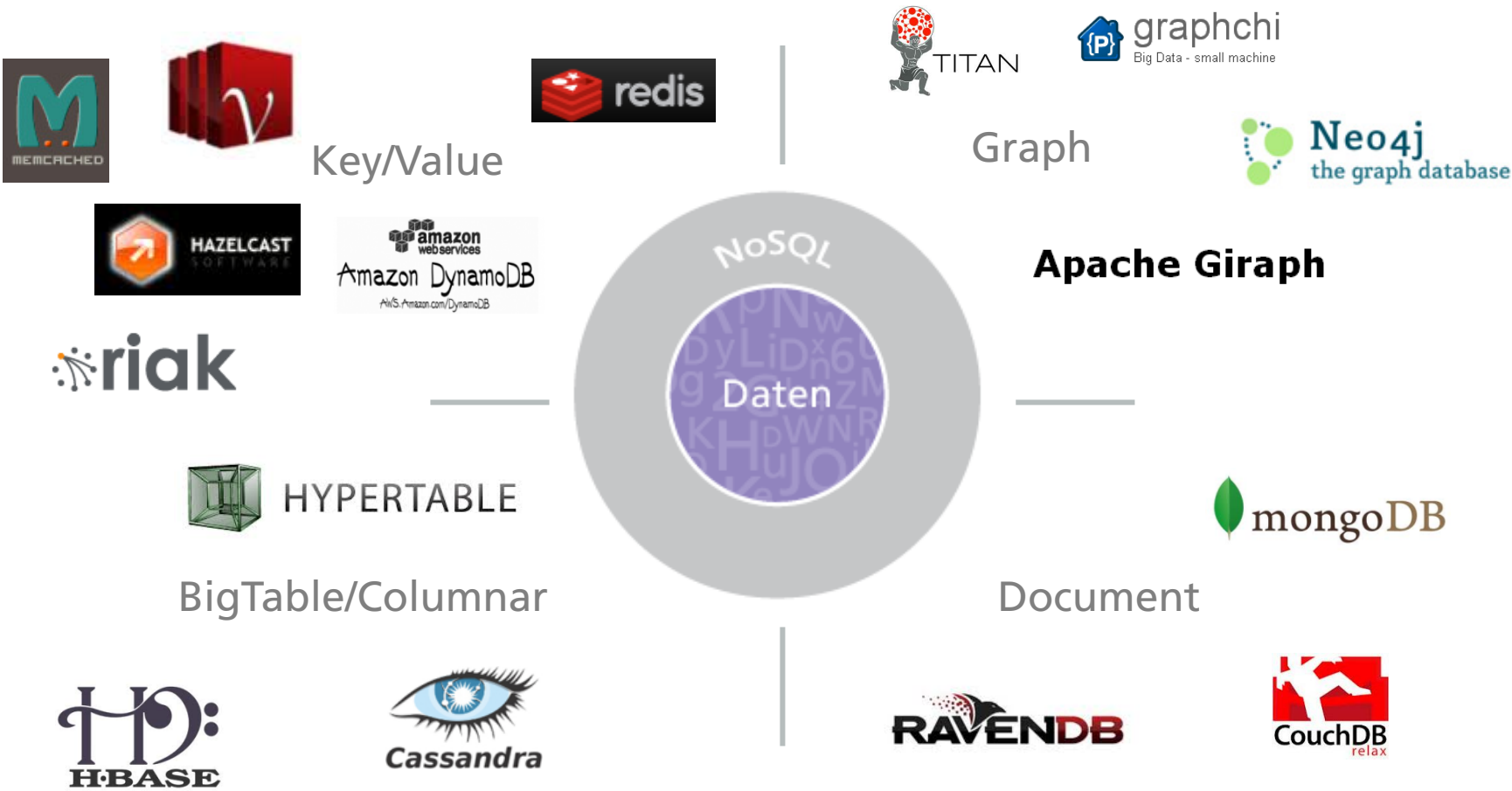
Im Kontext von Big Data sind bereits viele Open-Source Projekte verfügbar (eine Auswahl)



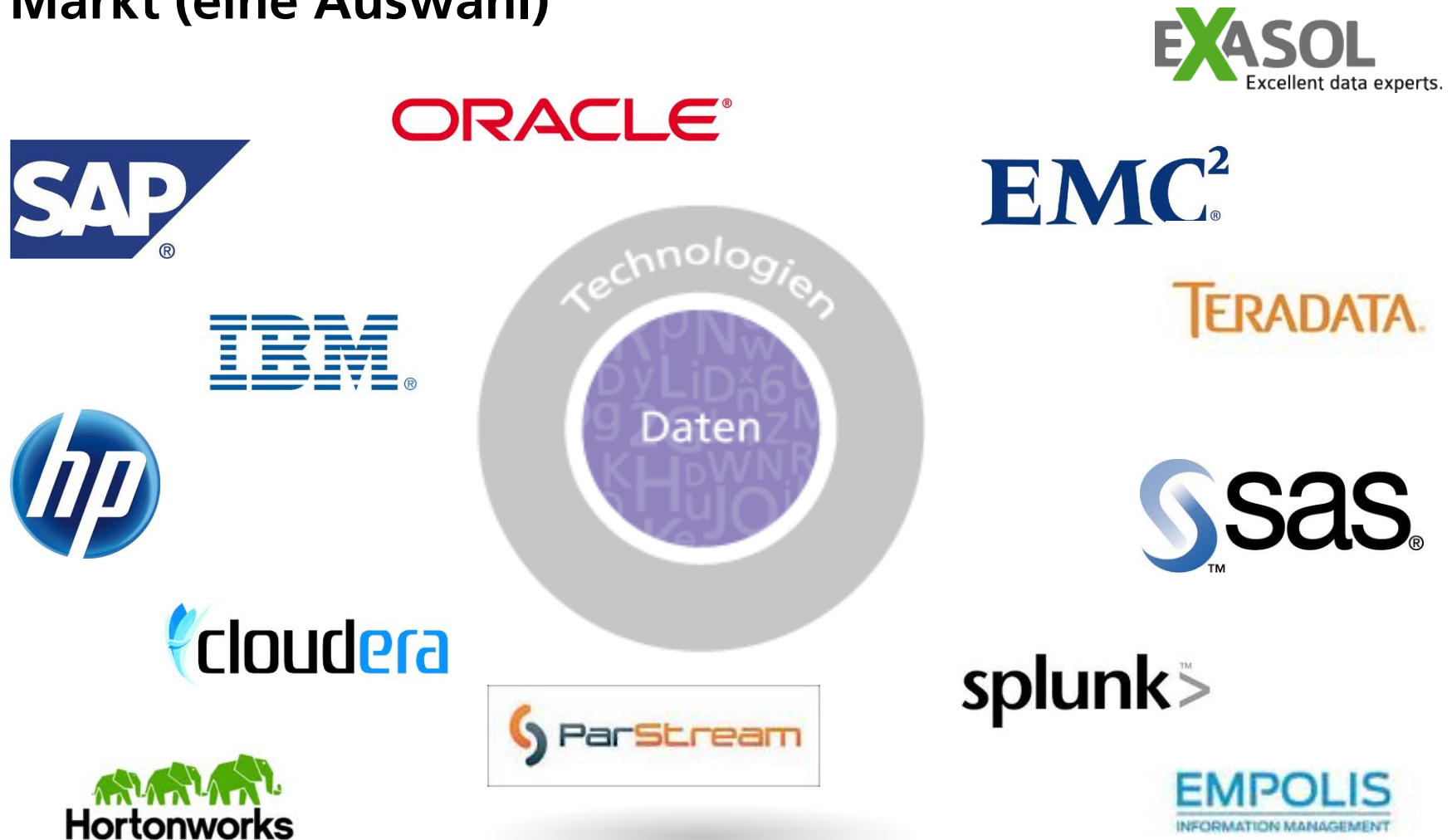
Big Data Anwendungen setzen auf eine Kombination von Tools aus verschiedenen Technologiekategorien



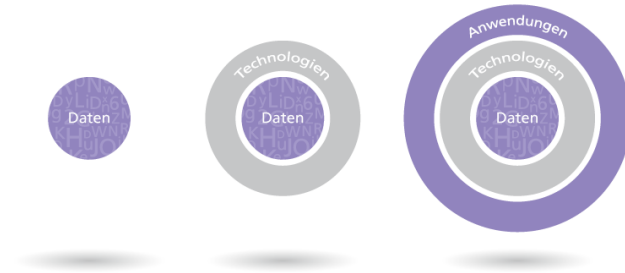
NoSQL Datenbanken – Verteilte nichtrelationale und schemafreie Datenbanken



Im Kontext von Big Data sind bereits viele Anbieter am Markt (eine Auswahl)

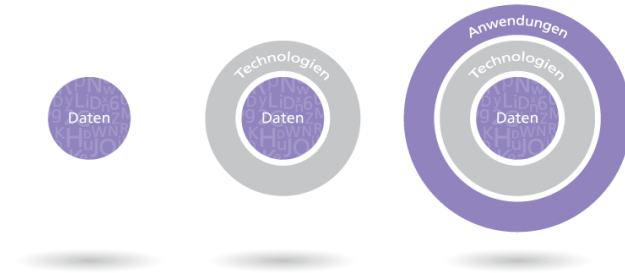


Barrieren



- Die Auswahl der richtigen Technologie ist eine Herausforderung
 - keine Standardisierung
 - unterschiedlicher Entwicklungsstand
 - unzureichende Dokumentation
 - uneinheitliche Terminologie, ...
 - Kaum Vergleichsmöglichkeiten, Benchmarks
- Kaum Personal mit Erfahrung vorhanden
- Anbieter reagieren mit „Out of the Box“-Lösungen (z.B. Appliance)
 - Hohe Kosten als Einstiegshürde

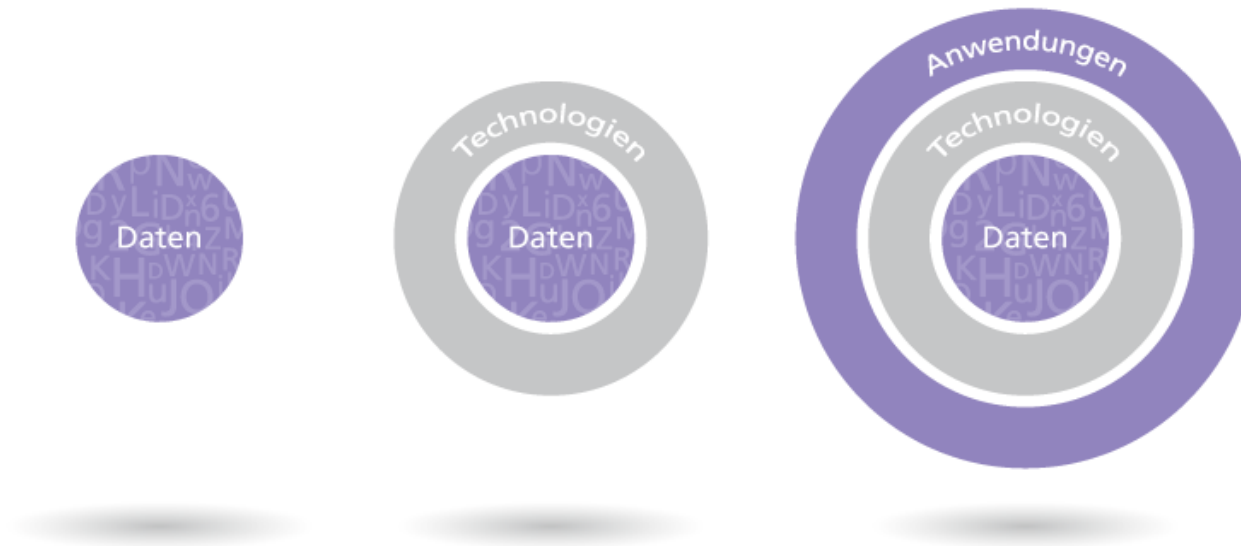
Ziele des Living Labs



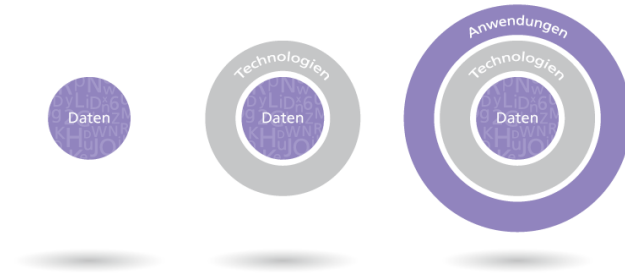
- Lern- und Experimentierumgebung für Unternehmen
- Technik „anfassbar“ machen, Einstiegshürden herabsetzen
- Teil eines Schulungskonzeptes „Data Science“
- Präsentation von Big-Data-Problemlösungen am Beispiel einer durchgängigen Anwendung des gesamten Technologie-Stacks
- Verschiedene öffentlich verfügbare Big-Data-Datenbestände werden eingespielt und stehen zum Testen zur Verfügung.
- Analyseverfahren für Big Data, die im Rahmen des THESEUS-Programms zum Thema Smart Semantics (Quote und Eat&Drink) entwickelt wurden
- Hardwareinfrastruktur steht nicht im Fokus
- Vorwettbewerblich, herstellerneutral, erweiterbar

Agenda

- Ziele des Living Lab Big Data
- **Living Lab Big Data Architektur**
- Living Lab Big Data Anwendungen
- Ausblick

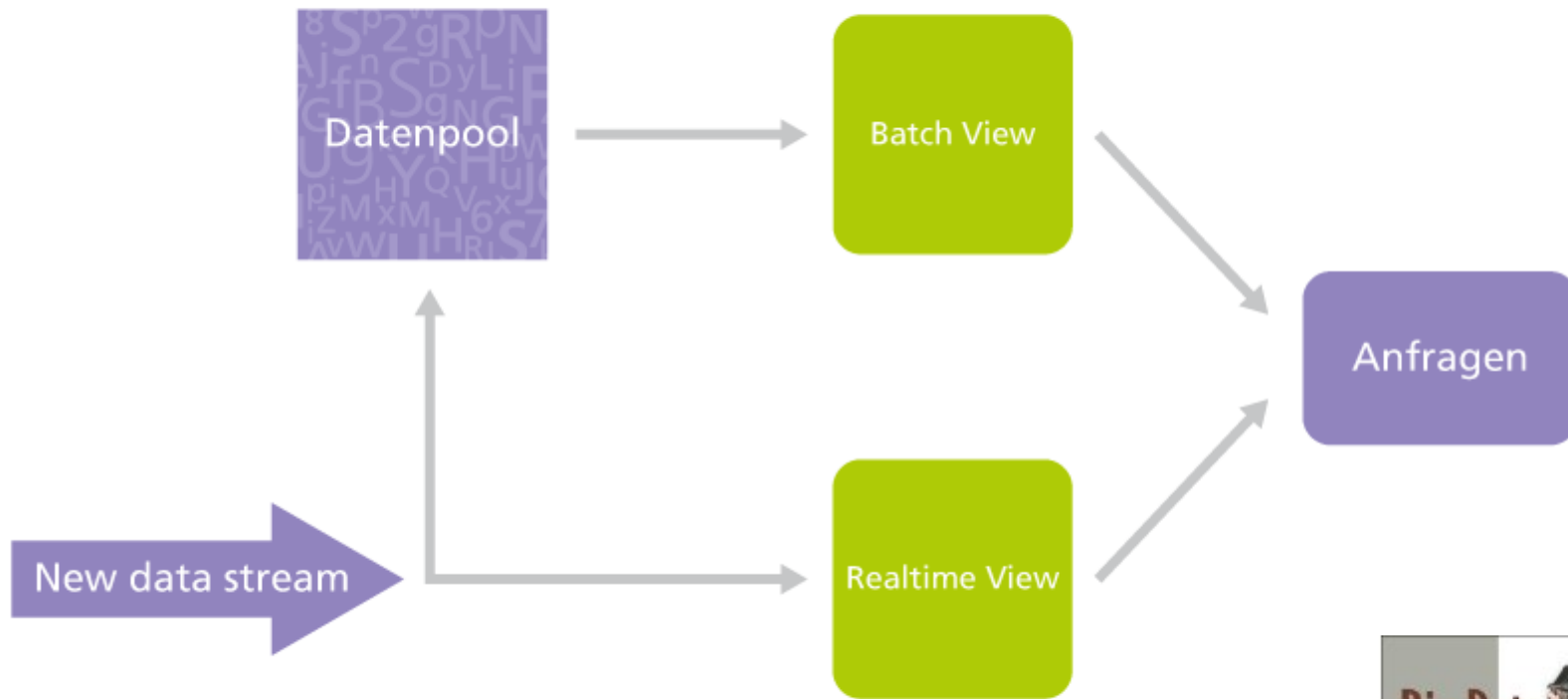


Design Living Lab



- Nicht eine einzelne Technologie (z.B. Hadoop, NoSQL), sondern eine **Big Data Architektur**, die flexibel genug ist,
 - verschiedenste Einsatzzwecke in verschiedensten Branchen abzudecken
 - es in den einzelnen Komponenten erlaubt, Technologien gegeneinander auszutauschen und z.B. auch eine individuelle Entscheidung zwischen OpenSource und kommerziellen Angeboten zu treffen
- Erst wenn man in Architekturen denkt, erschließen sich Stärken und Schwächen existierender Technologien

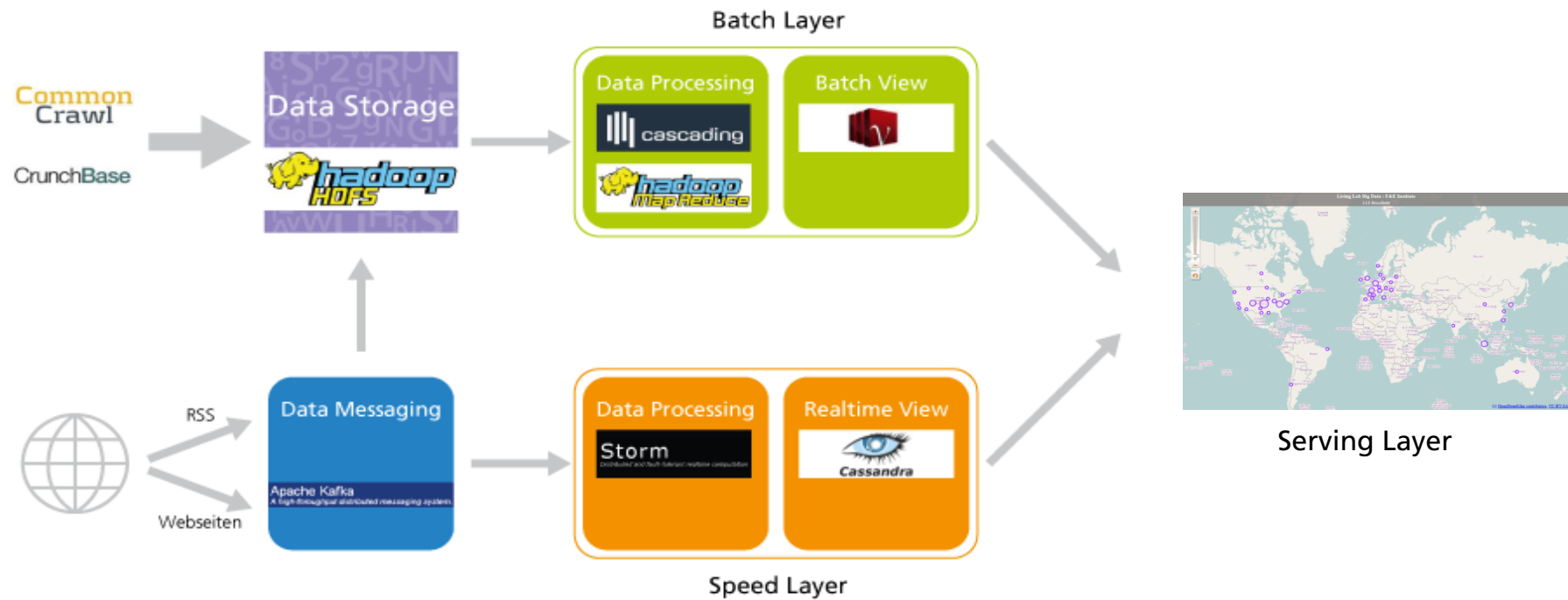
«Lambda Architecture»



Quelle: Nathan Marz. Big Data: Principles and Best Practices of Scalable Realtime Data Systems. 2013

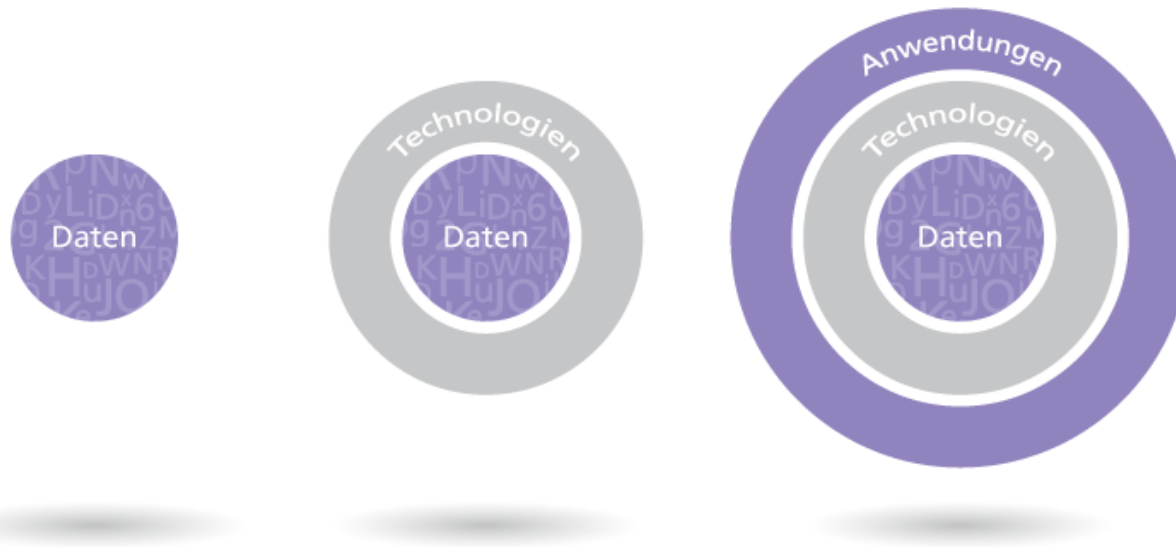


Living Lab - Architektur



Agenda

- Ziele des Living Lab Big Data
- Living Lab Big Data Architektur
- **Living Lab Big Data Anwendungen**
- Ausblick



Big Data Experimentierplattform – Technologie zum Anfassen im Rahmen des Schulungsmoduls „Data Scientist Big Data“



Batch-Anwendung
Analyse von
Kundenfeedback

Realtime-Anwendung
Big Data
Themenmonitor

Common
Crawl

6 Milliarden
Webseiten
(Q1/2012)
~ 20TB nur Text

Ausgewählte
Technologien

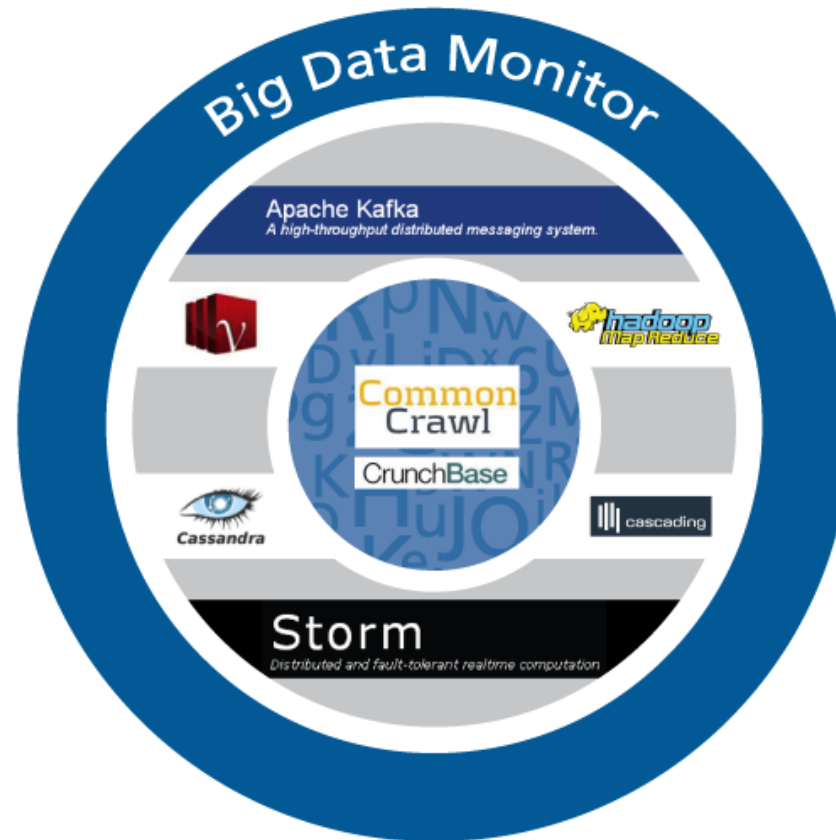


Anwendungs-
fälle

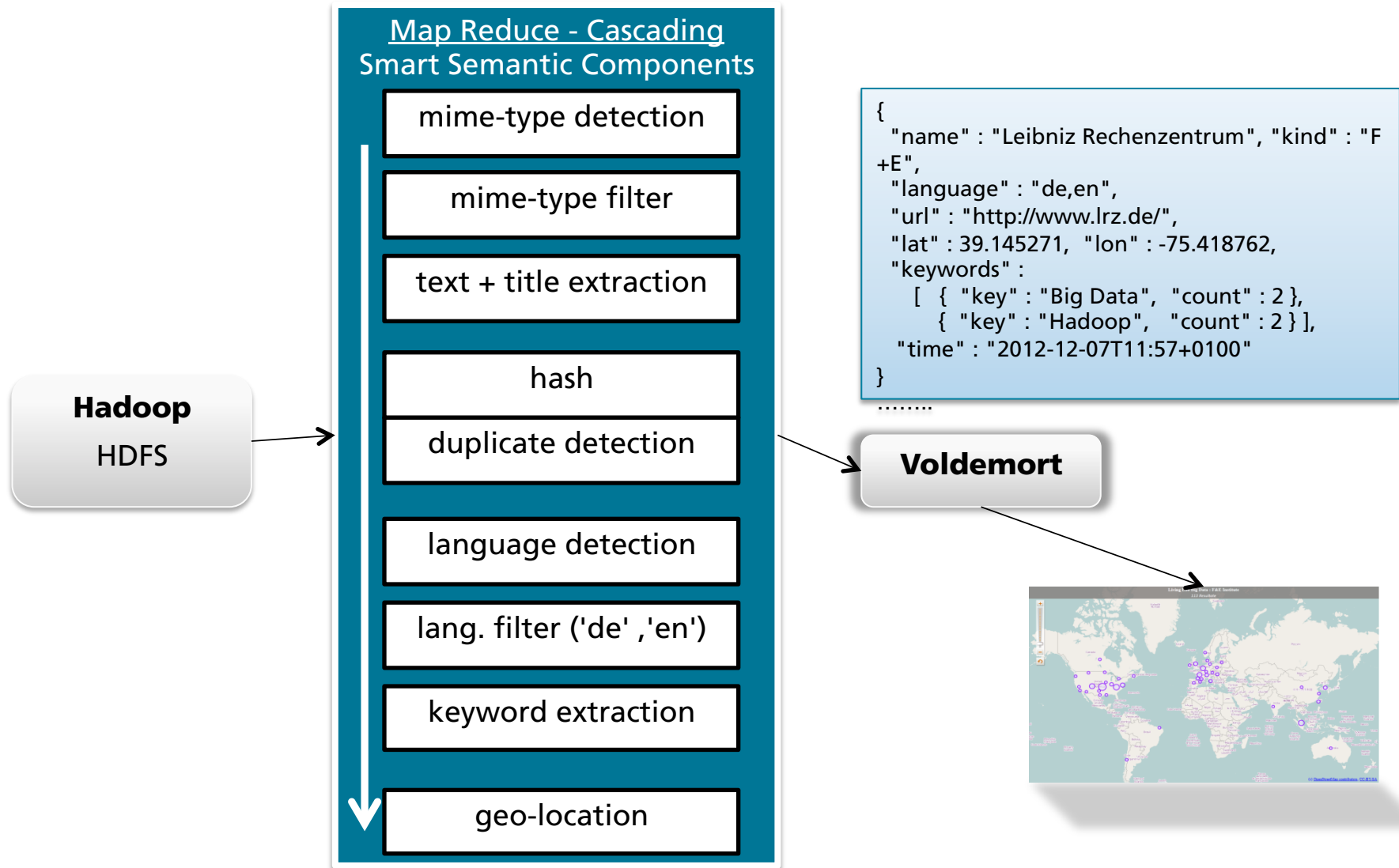


Big Data
Datensatz

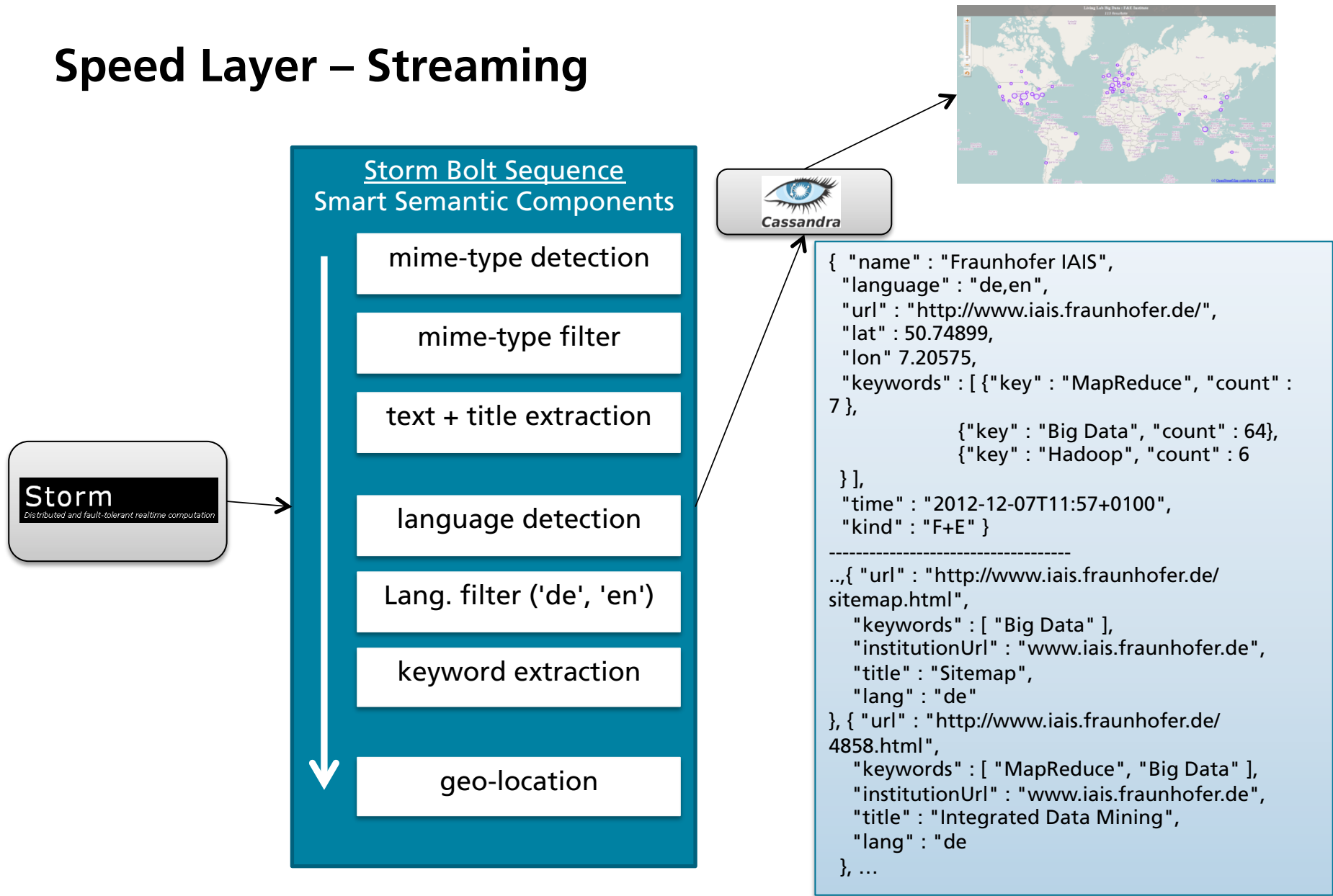
Anwendungsfall: Monitoring von Themen



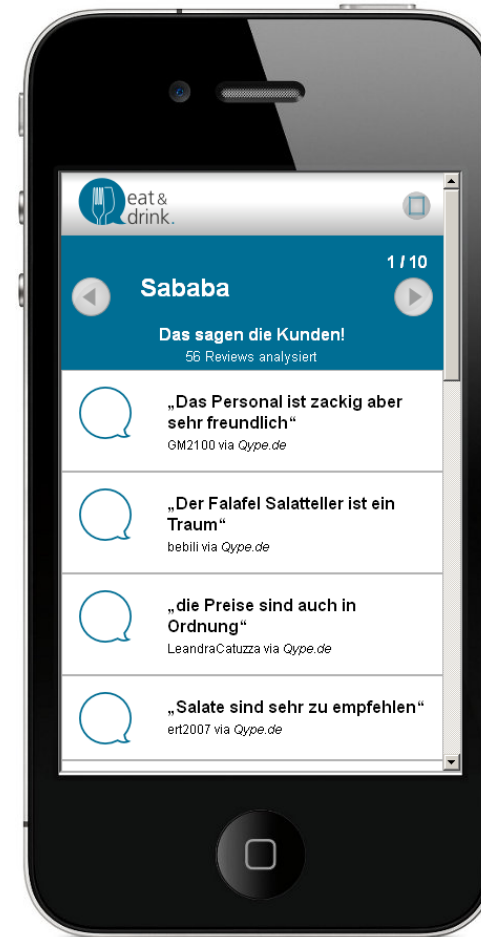
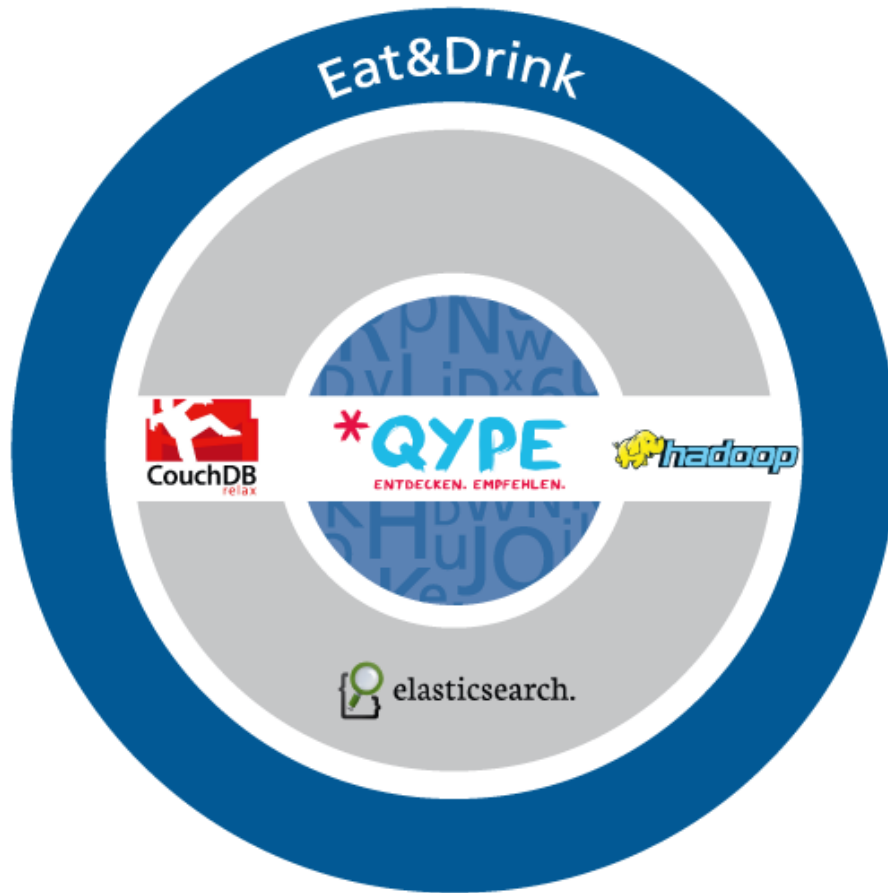
Batch Layer - Zyklus von Stunden oder Tagen



Speed Layer – Streaming

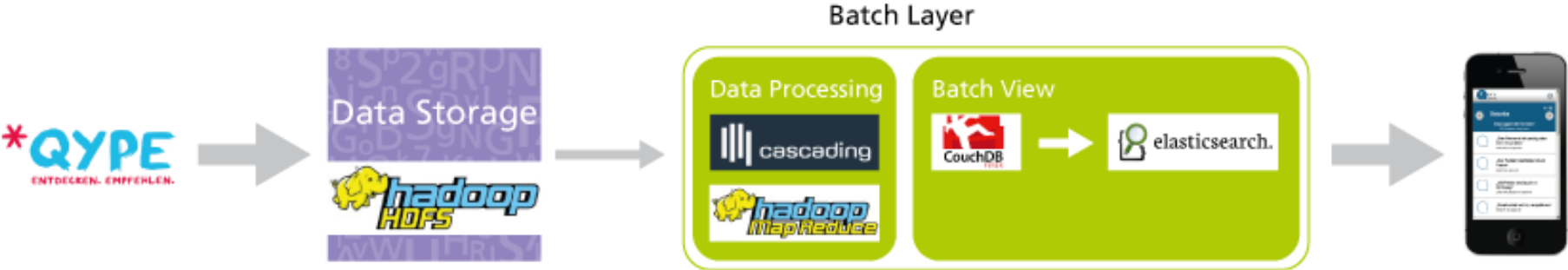


Anwendungsfall: Kundenfeedbackanalyse

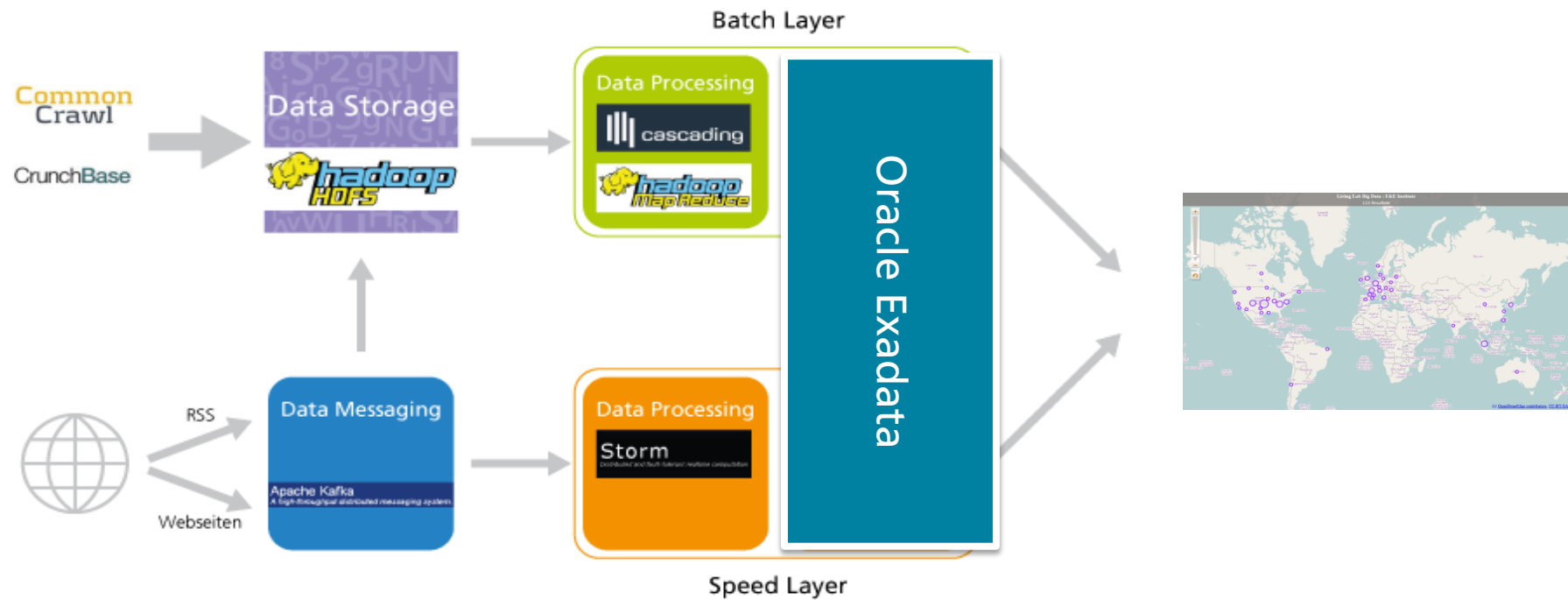


App Eat&Drink

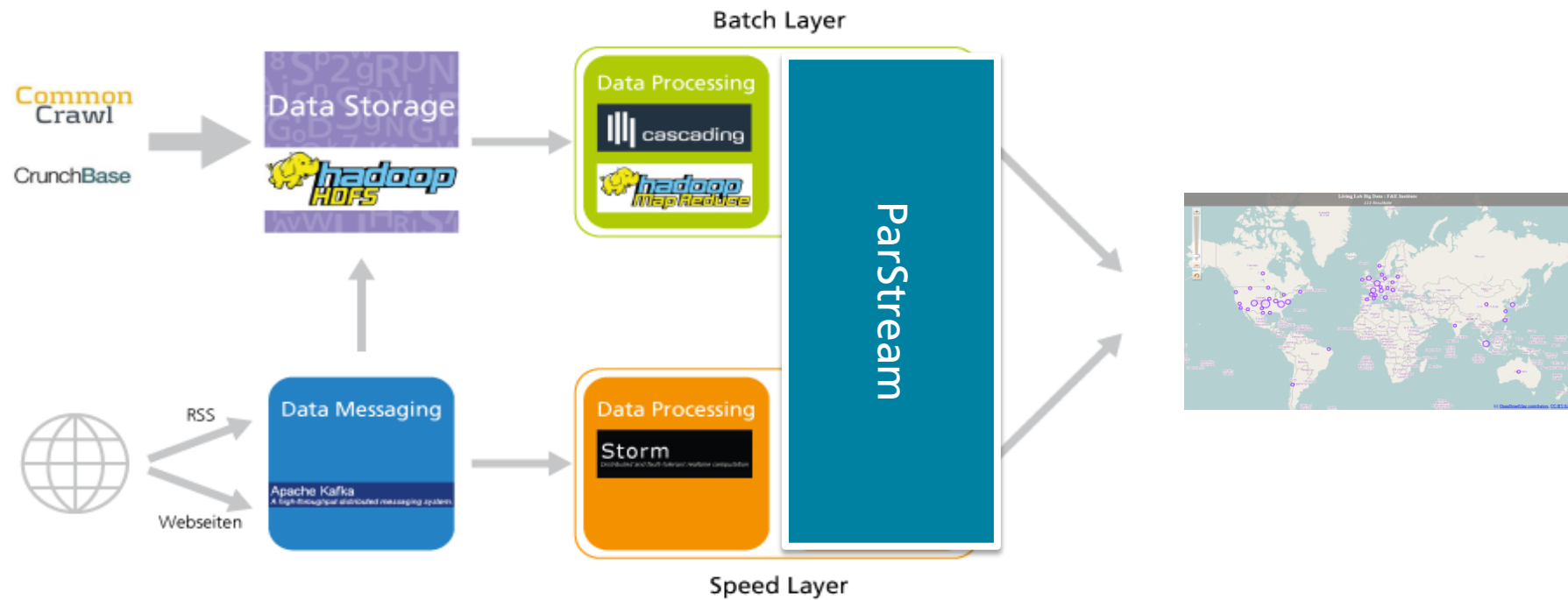
Living Lab Use Case 2 – Workflow



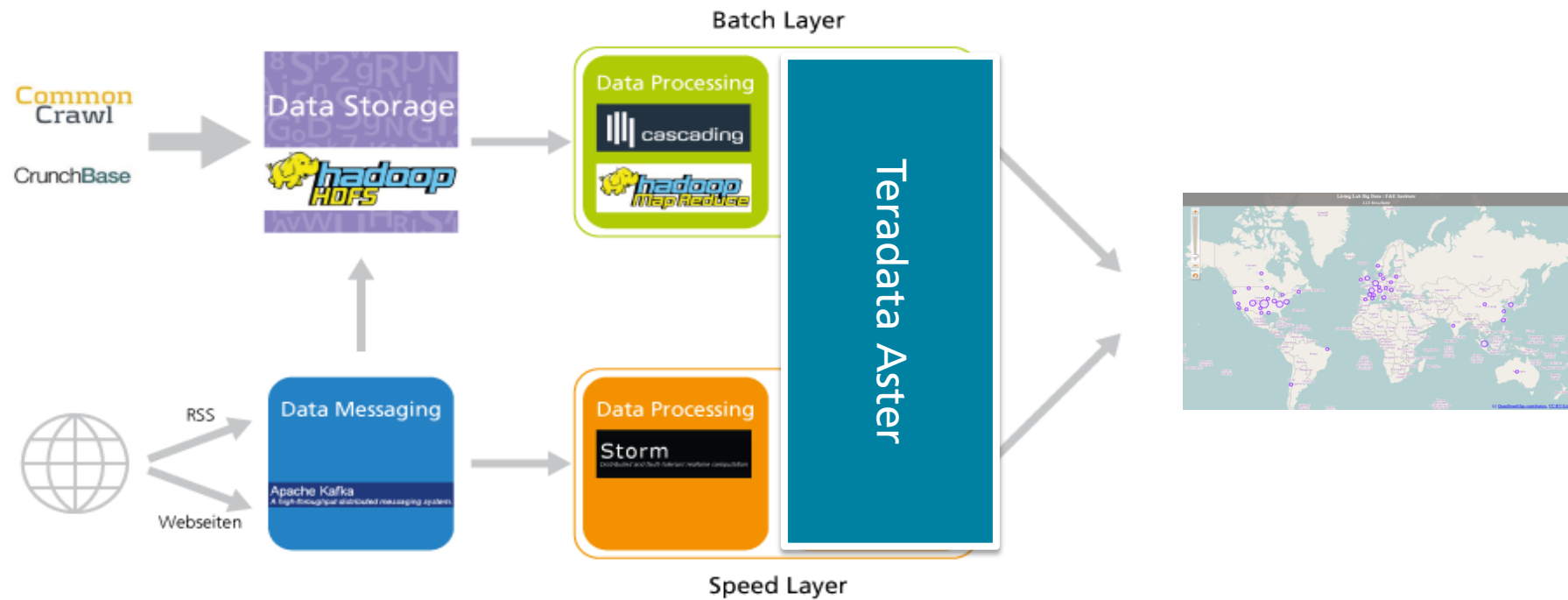
Living Lab Technologiealternativen



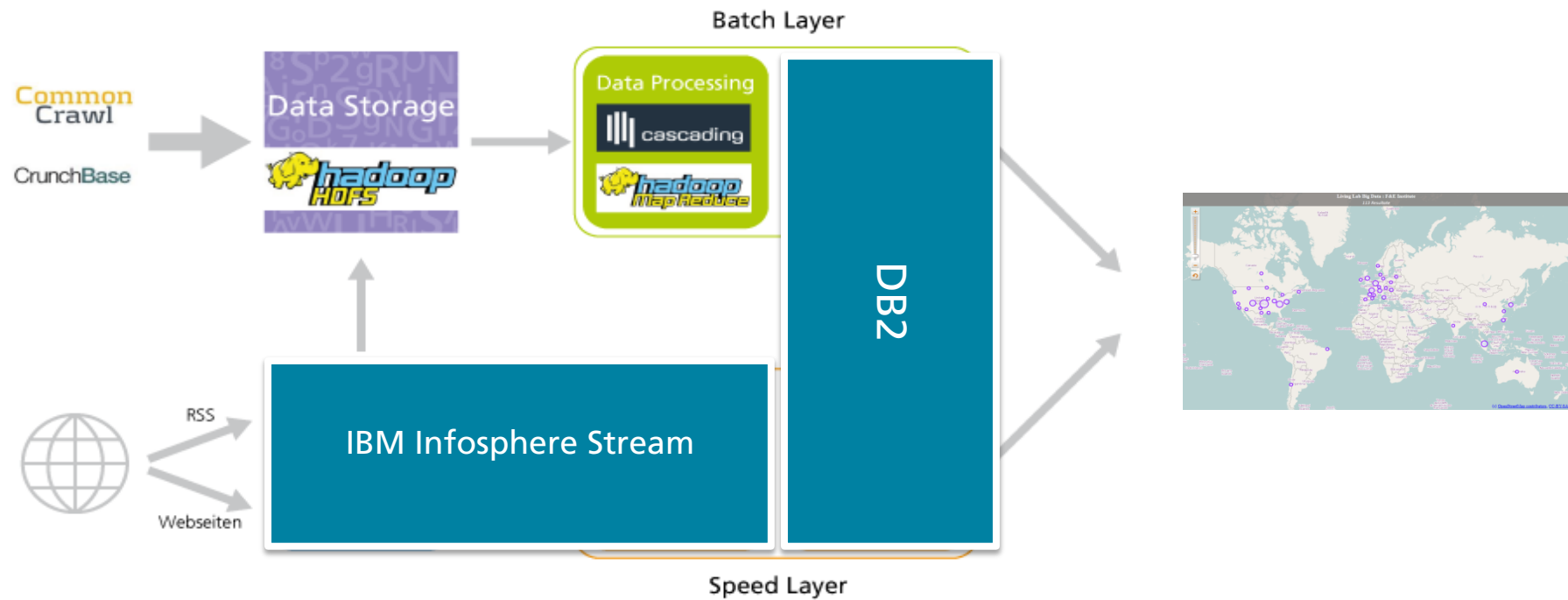
Living Lab Technologiealternativen



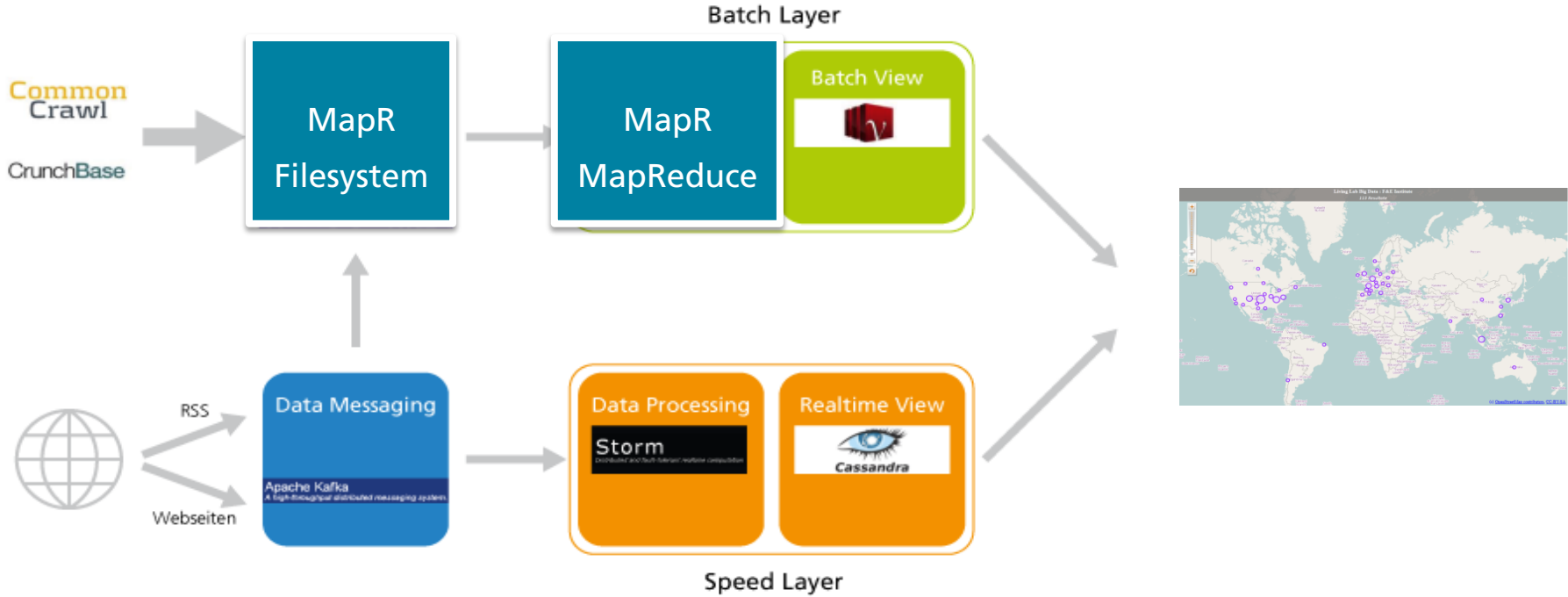
Living Lab Technologiealternativen



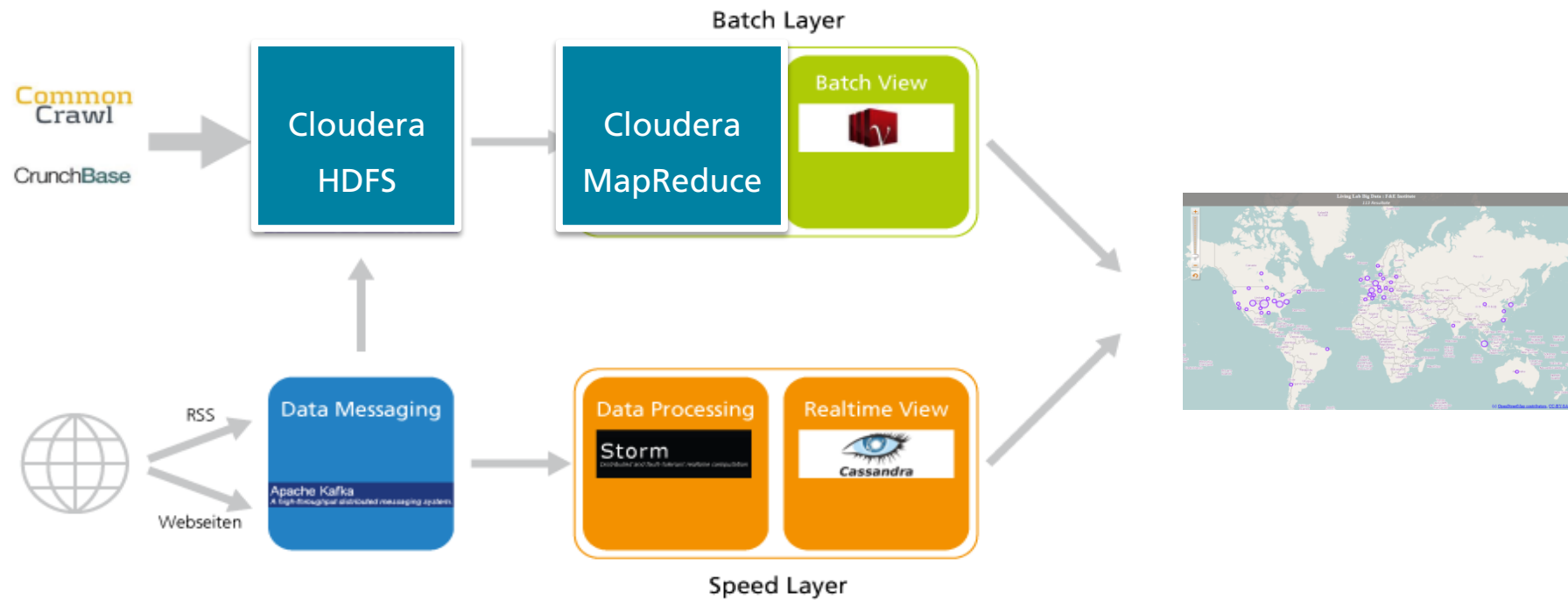
Living Lab Technologiealternativen



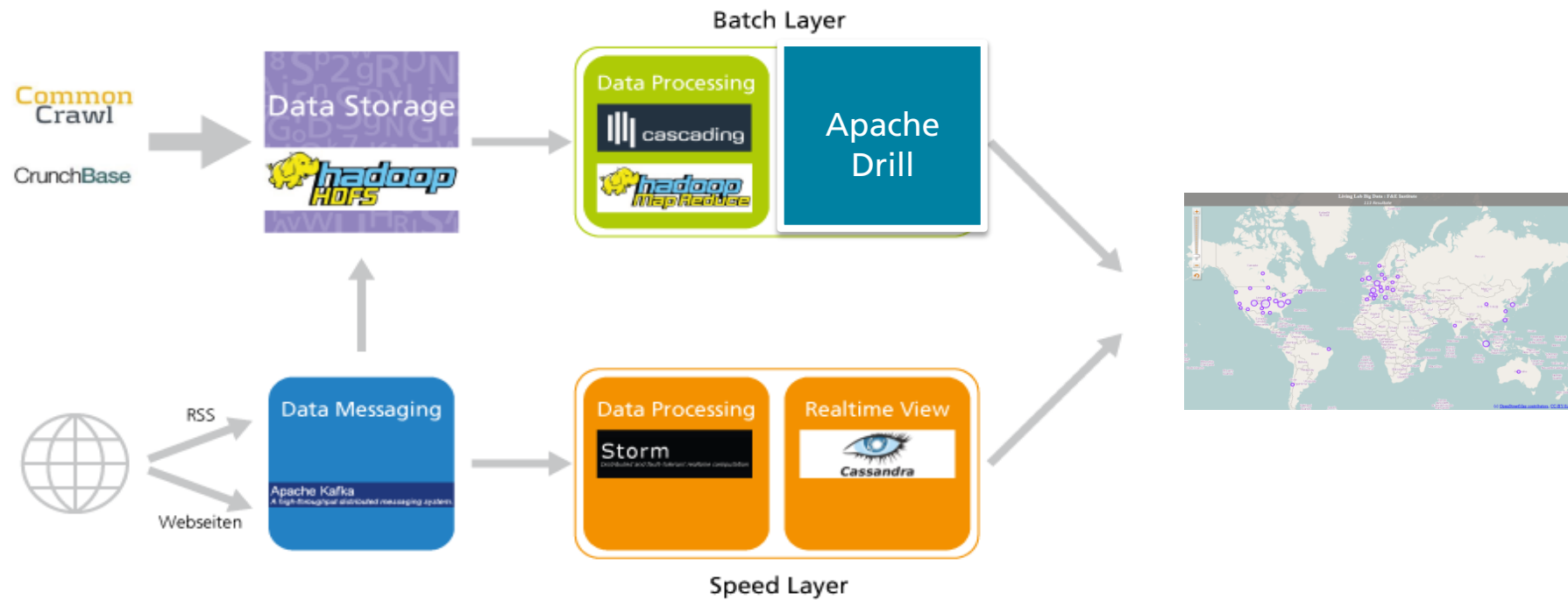
Living Lab Technologiealternativen



Living Lab Technologiealternativen

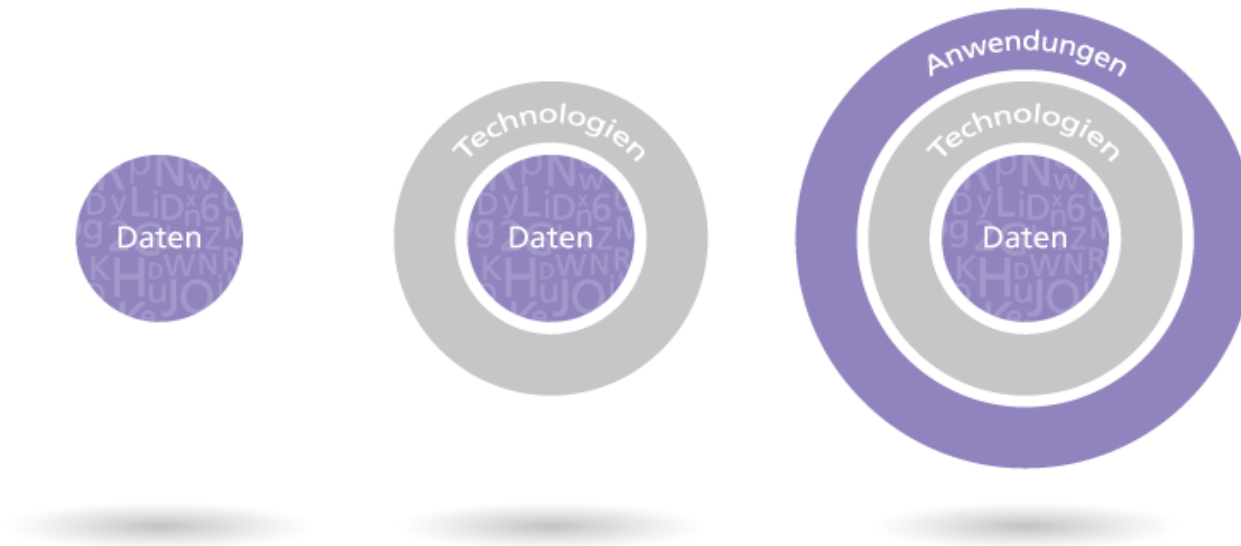


Living Lab Technologiealternativen



Agenda

- Ziele des Living Lab Big Data
- Living Lab Big Data Architektur
- Living Lab Big Data Anwendungen
- **Ausblick**



Wie kann das Living Lab genutzt werden?

- Zielgruppe: Data Scientists, Analysten, Entwickler
- Teil des Schulungskonzeptes Data Scientist – „Big Data Management“
- Schulungsmodul
- Individuelle Workshops
- „Bring your own Data“
- In-House-Installation möglich

Geplante Data Scientist Schulungen von Fraunhofer IAIS

Big Data
Management

2-3 Tage

Ab März 2013

Data Scientist
Natural
Language
Processing
(NLP)

2 Tage

Ab Oktober 2012

Data Scientist
Big Data
Analytics
Basics

2 Tage

Ab Februar 2013

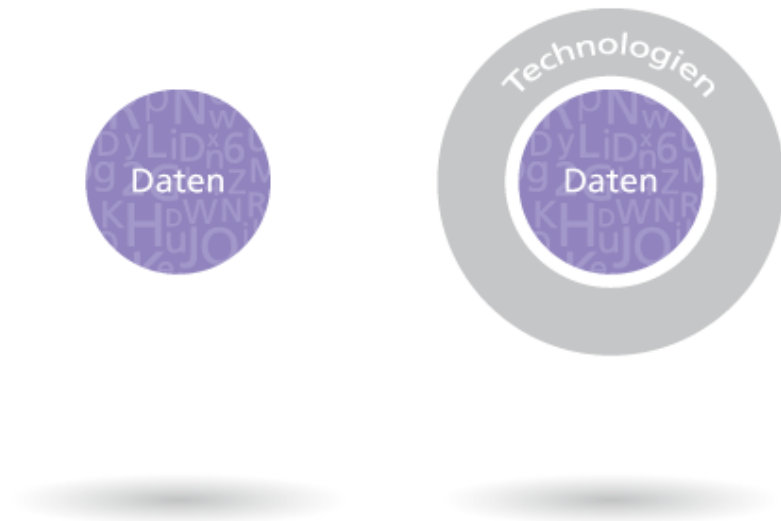
Data Scientist
Visual
Analytics

2 Tage

Ab Februar 2013

Nächste Schritte

- Diskussion von Referenzarchitekturen im Arbeitskreis Big Data der BITKOM
- Gespräche mit Anbietern
- Public-Private-Partnership
- Living Lab Big Data@CEBIT 2013



Zusammenfassung

- Das Living Lab macht Big Data „anfassbar“
- Ziel: Orientierung, Best Practices, Schulung
- Zentrales Architekturkonzept: Lambda Architektur
 - Basiert auf Open-Source-Komponenten
 - Durch kommerzielle Komponenten teilweise substituierbar
 - Erlaubt Exploration von Alternativen
- Herstellerunabhängig, vorwettbewerblich

