



Hochschule für Angewandte Wissenschaften Hamburg

Hamburg University of Applied Sciences

ANWENDUNGSSPEZIFISCHE PROAKTIVE BEHANDLUNG VON PROZESSEN IM BEREICH BIG DATA

Lotaire Tchamadeu Tiappi
Angewandte Informatik HAW
Hamburg, Deutschland
Lotaire.tchamadeutiappi@haw-hamburg.de



Anwendungsspezifische proaktive Behandlung von Prozessen im Bereich Big Data
Im Rahmen des Grundseminars
im Studiengang Master Informatik
am Department Informatik
der Fakultät Technik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuer: Prof Dr. Kai von Luck und Prof. Dr. Tim.Tiedemann
Abgabe am . 18.04.2017

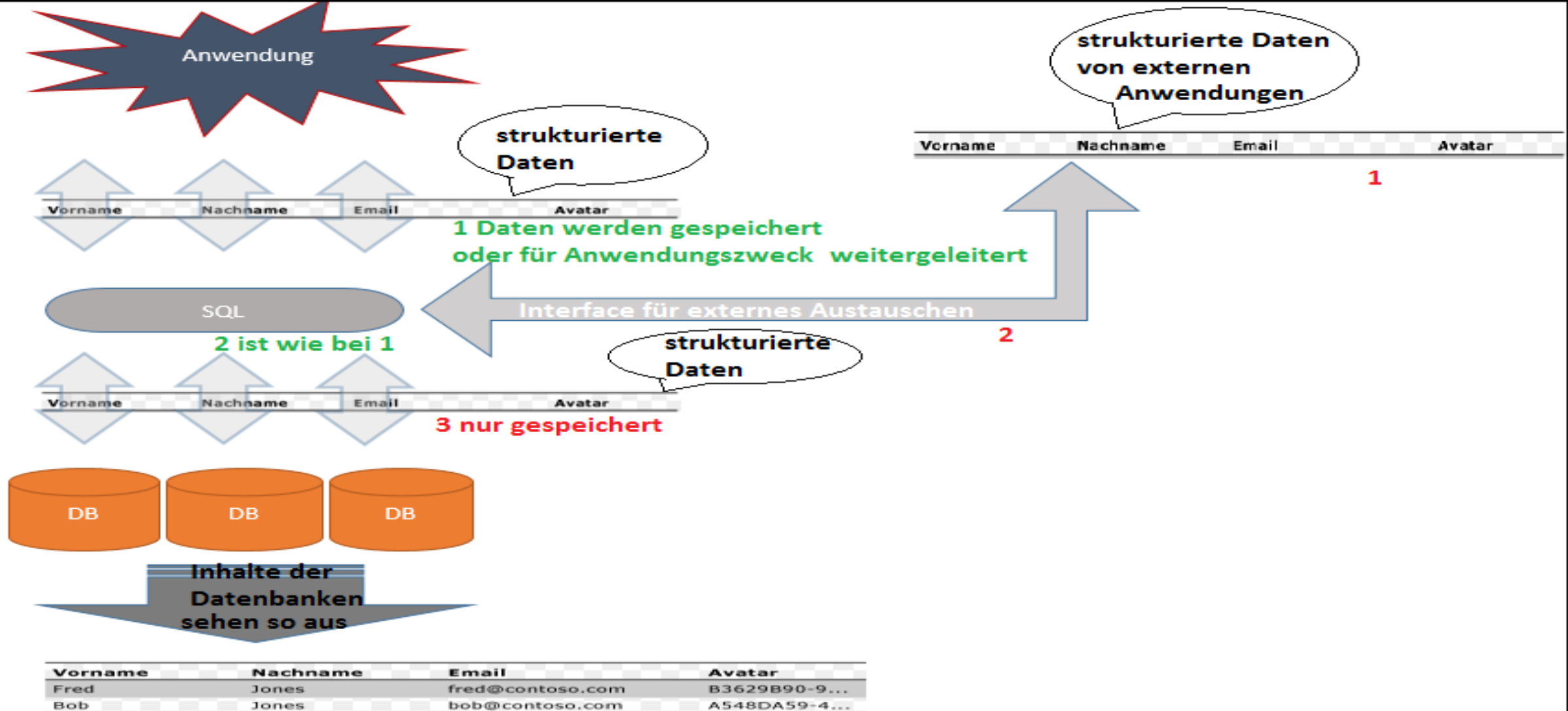
GLIEDERUNG

- **KLASSISCHES DATENMODELLVERFAHREN**
- **EREIGNISSE ZUM BIG DATA**
- **BIG DATA SOFTWARE ENGINEERING**
- **ANWENDUNGSGEBIET**
- **ZIEL DES FORSCHUNGSBEREICHES**
- **FAZIT**

KLASSISCHES DATENMODELLVERFAHREN

KLASSISCHES DATENMODELLVERFAHREN

- **STRUKTURIERTE DATEN**
- **SQL ABFRAGE**
- **DATEN KOMMUNIZIEREN ÜBER SCHNITTSTELLE ...**
- **AGGREGIERTE ZUGRIFFE AUF DATEN MÖGLICH**



Klassisches Datenmodellverfahren Eigene Darstellung

EREIGNISSE ZUM BIG DATA

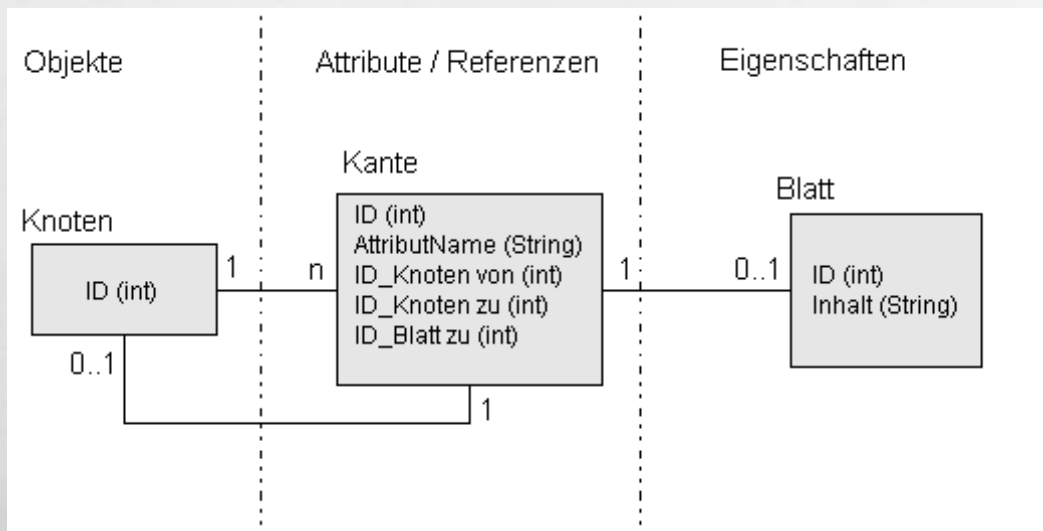
EREIGNISSE ZUM BIG DATA



- **VERNETZUNG**
- **IOT**
- **DIGITALISIERUNG**
- **INDUSTRIE. 4.0**

EREIGNIS ZUM BIG DATA

NEUE ARTEN VON DATEN



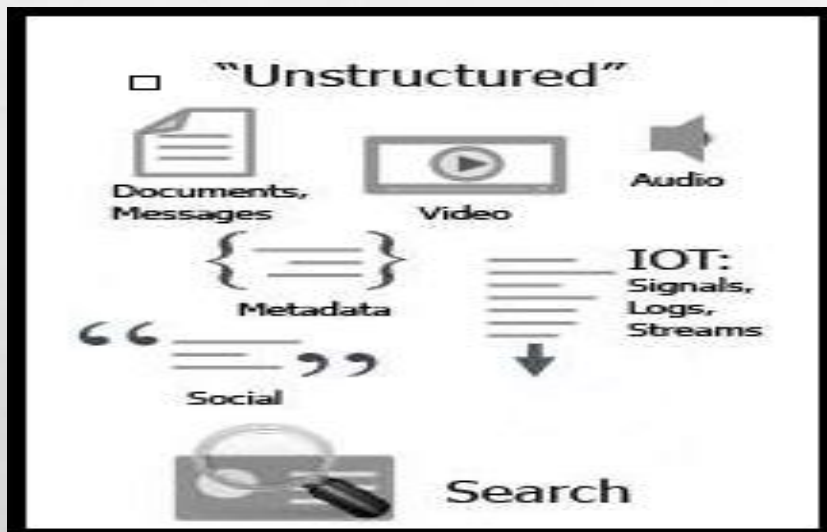
Relationales Datenmodell zur Abbildung von semistrukturierte Objekten!1

SEMI-STRUKTURIERTE DATEN

- **EINE ODER MEHRERE ARTEN VON OBJEKTEN**
- **OBJEKT ALS ATTRIBUT ODER OTOMARE**
- **OBJEKTE ENTHALTEN WERTE EINES BEKANNTEN ELEMENTAREN DATENTYPS**

EREIGNIS ZUM BIG DATA

NEUE ARTEN VON DATEN



Unstrukturierte Daten[2]

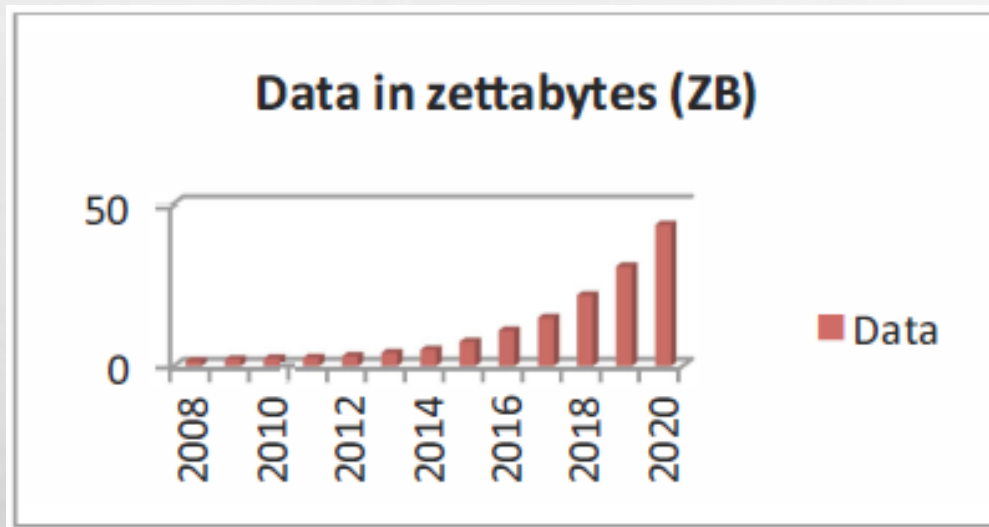
UNSTRUKTURIERTE DATEN

LÖSUNG FÜR DIE STRUKTURIERUNG DER DATEN

- **MASCHINELLES LERNEN**
- **TEXTANALYSE**
- **LINGUISTISCHE VERFAHREN**

EREIGNIS ZUM BIG DATA

ZUKÜNFTIGE PROGNOSE

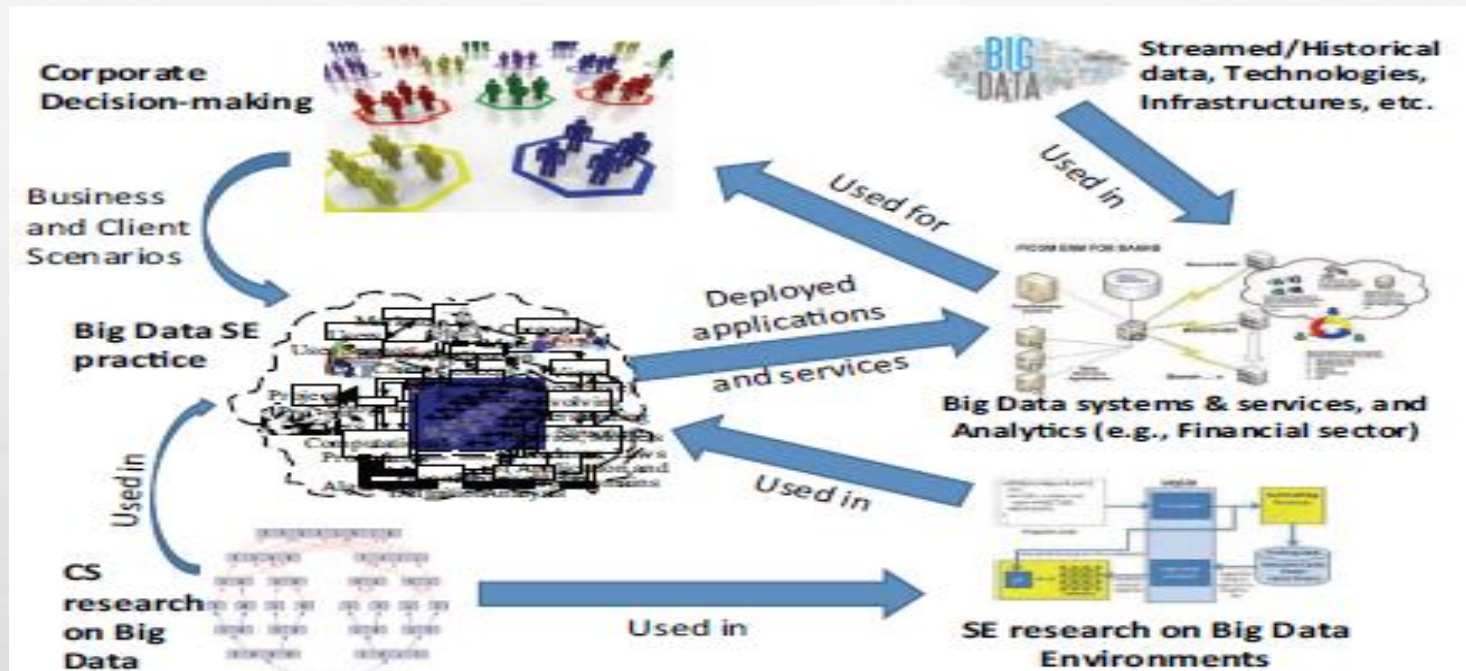


Oracle Quelle 2012[3]

Daten werden immer mehr jedes Jahr

BIG DATA SOFTWARE ENGINEERING

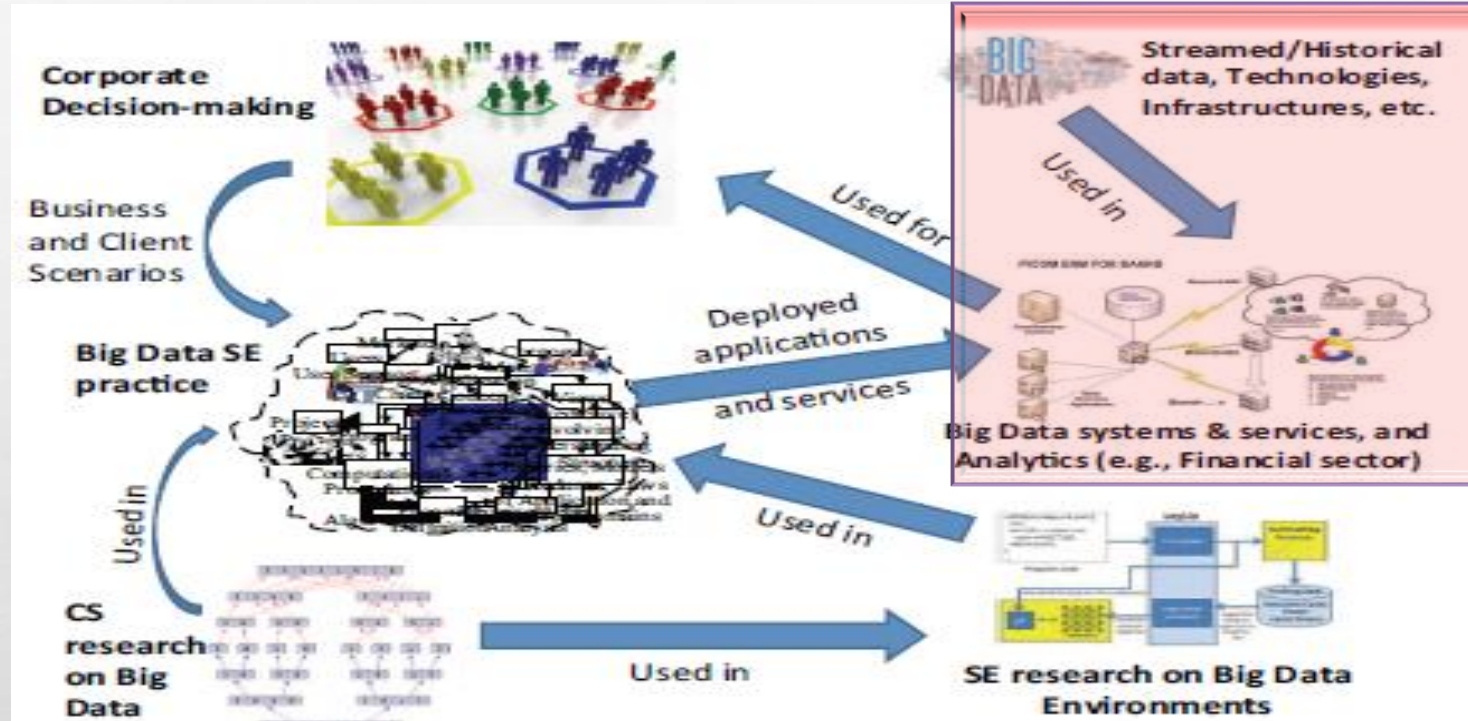
BIG DATA SOFTWARE ENGINEERING



Context Model of Big Data Software (4)

- **Unser Thema fokussiert sich Im Bereich Big Data Engineering .**
- **Das Bild gibt Hinweise auf die breiten Ansätze und die Bedeutung dieses Entwicklungsgebiets für den aktuellen Wettbewerb**

BIG DATA SOFTWARE ENGINEERING



Context Model of Big Data Software [4]

- **Spezifische Anwendungen und Big Data Bereich**
- **Proaktive Behandlungen von Prozessen**
- **Frameworks spezifische Ansätze für definierte Behandlung**

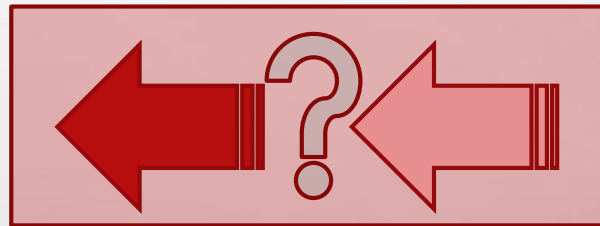
STREAM VERFAHREN



BIG DATA

Daten entstehen am meisten aus mehreren Quellen. Diese können aus strukturierten oder unstrukturierten Daten bestehen.

Das Datenspeichern wird durch vielfältige Komponenten unterstützt.



**Data Quelle:
Sensor, GPS...**

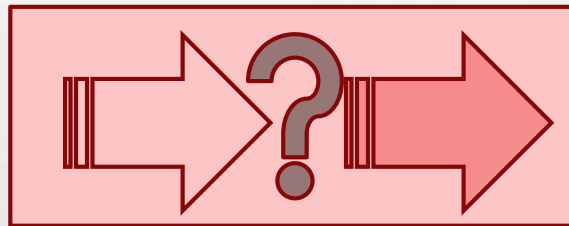
Quelle: Eigene Darstellung

STREAM VERFAHREN

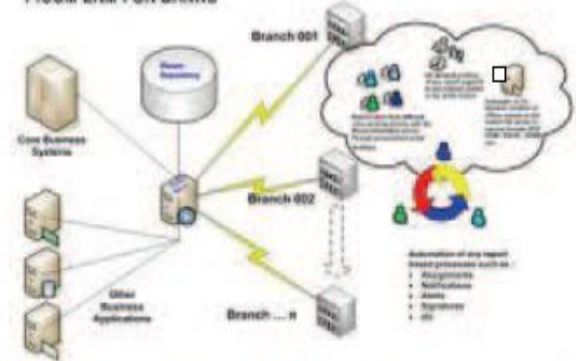
Auch die Bereitstellung (Datenzugriff) von Daten für spezifische Applikationen werden durch mehrere Frameworks unterstützt.



Streamed/Historical data, Technologies, Infrastructures, etc.



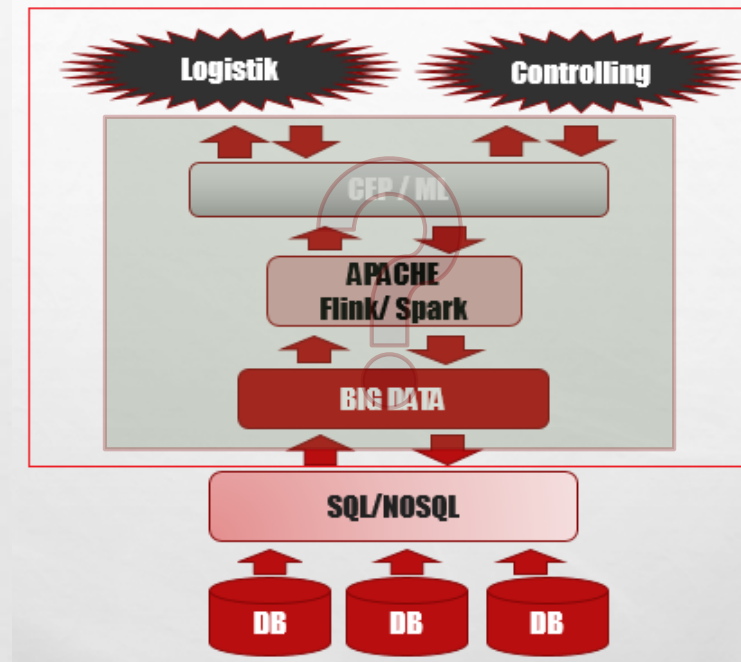
PICOM ERM FOR BANKS



Big Data systems & services, and Analytics (e.g., Financial sector)

Quelle: Eigene Darstellung

PROAKTIVE LÖSUNGEN



Bestandteil des Austauschverfahrens (eigene Darstellung)

- **CEP/ ML**
- **Apache FLINK**
- **Apache SPARK**
- **Apache HADOOP**
- **Apark + Hadoop**

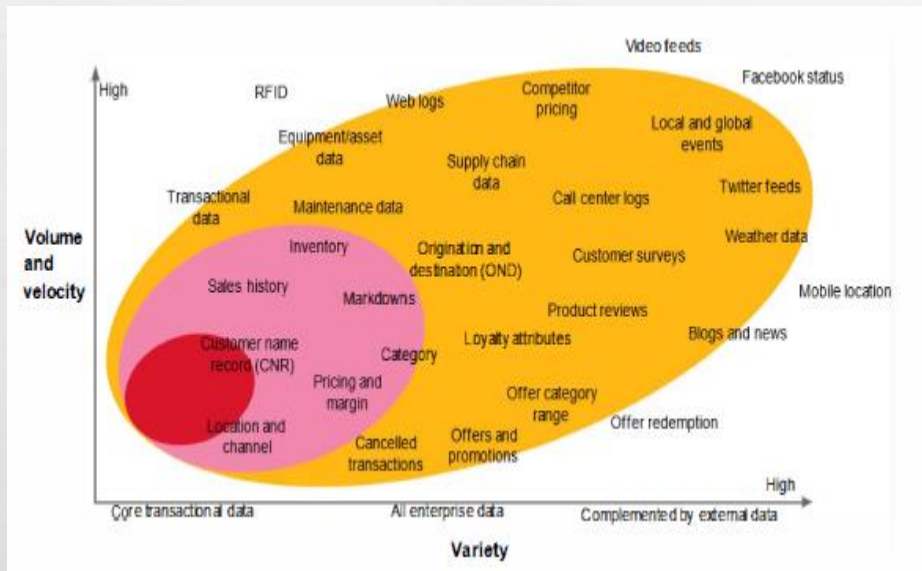
ANWENDUNGSGEBIET

ANWENDUNGSGEBIET

- **LOGISTIK**
- **SUPPLY CHAIN MANAGEMENT**
- **CONTROLLING ...**

ANWENDUNGSGEBIET

LOGISTIK



**Unendliche Anzahl von Datenquellen;
Auto-Sensor, auf der Straße,
Gps, Kundenanwendungen ...**
**Analyse Lösung
- Apache Hadoop**

Large variety of travel and transportation companies data,p.3[5]

HADOOP IM LOGISTIK

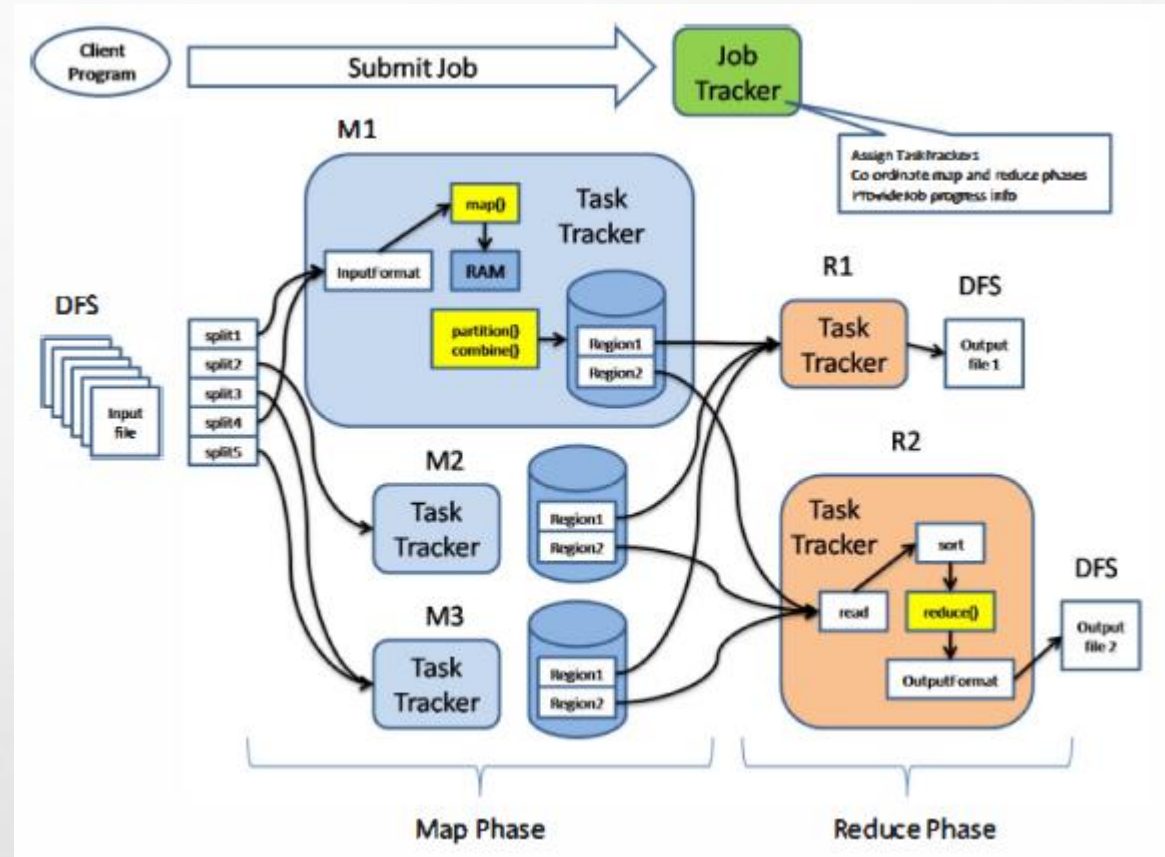
LESEN VON CONTAINERBESCHRIFTUNGEN DURCH
MAP REDUCE

. MAP ETAPE

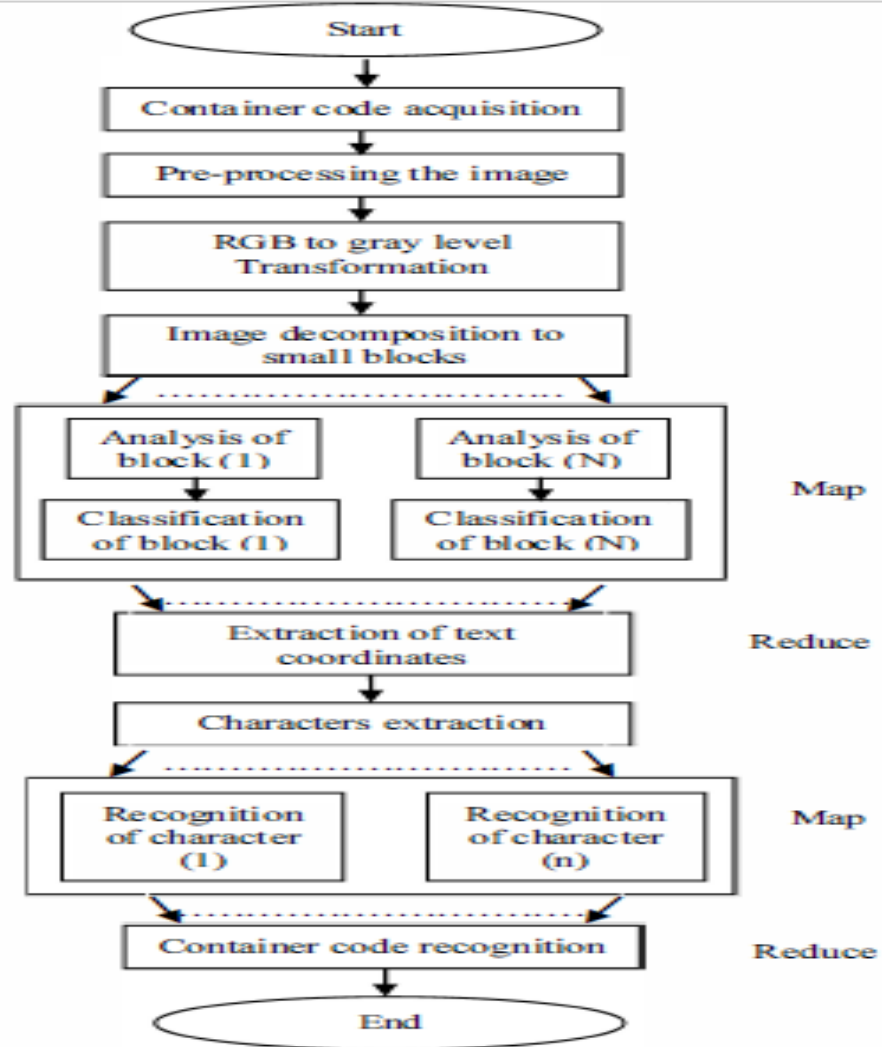
1 BESCHRIFTUNG SCANNEN, ZERLEGEN, VERTEILEN AN MEHRERE
STACK

. REDUCE ETAPE

BUCHSTABEN ZUSAMMENBRINGEN UND DIE INFO EXTRAHIEREN



MapReduce architecture [6]



Big data System für Container code Erkennung(Shyam Ra*.u.a. 2015.)

ANWENDUNGSGEBIET

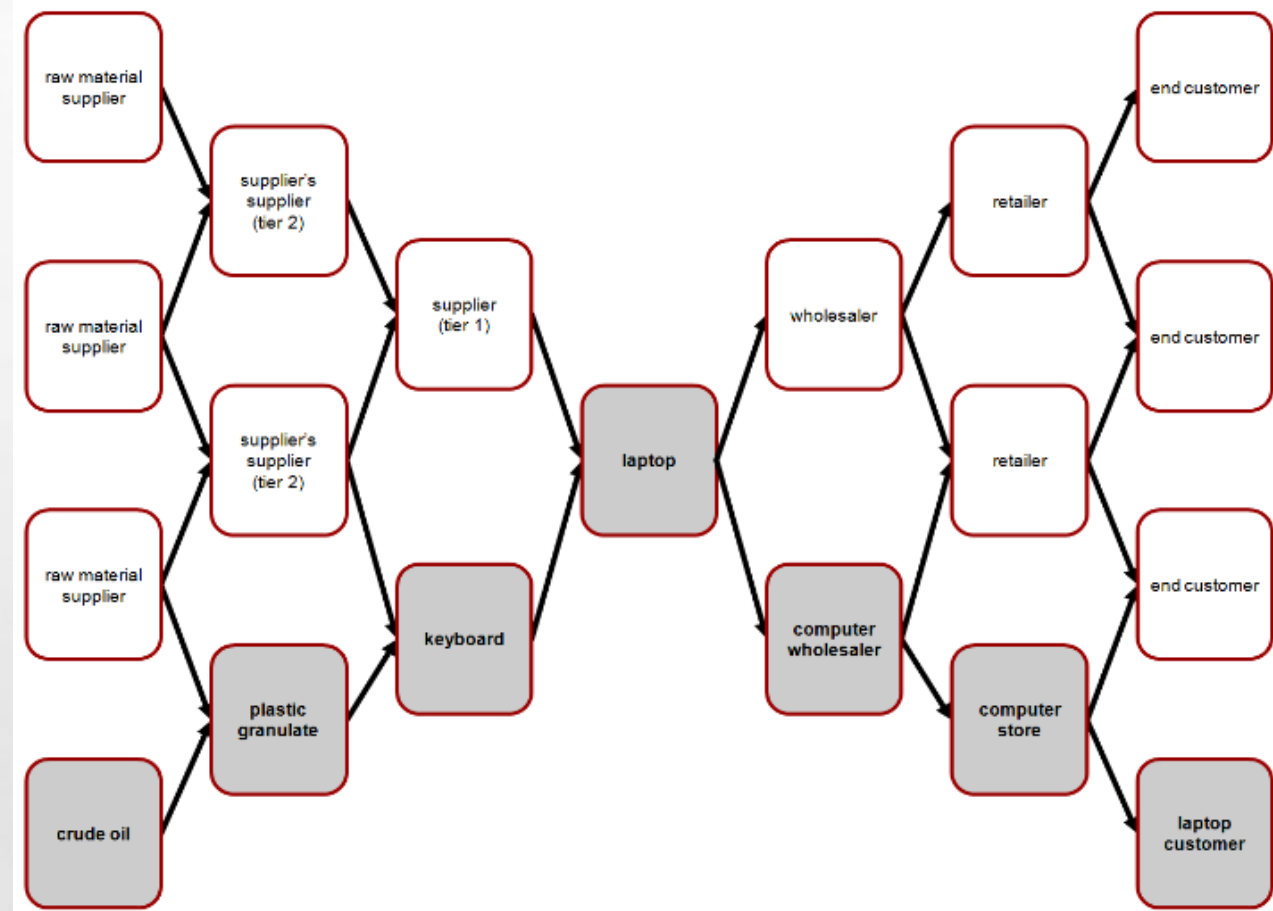
SUPPLY CHAIN MANAGEMENT

DIE HAUPTAUFGABE DES LIEFERANT-MANAGEMENTS IST ES, DAS LIEFERUNGSRISIKO BEI DEN LIEFERANTEN FRÜH ZU ERKENNEN UND ZU BEKÄMPFEN.

AUFGABE DER SCM UND APACHE SPARK IM KONTEXT DER SCM

SUPPLY CHAIN MANAGEMENT UND EINIGE AUFGABEN

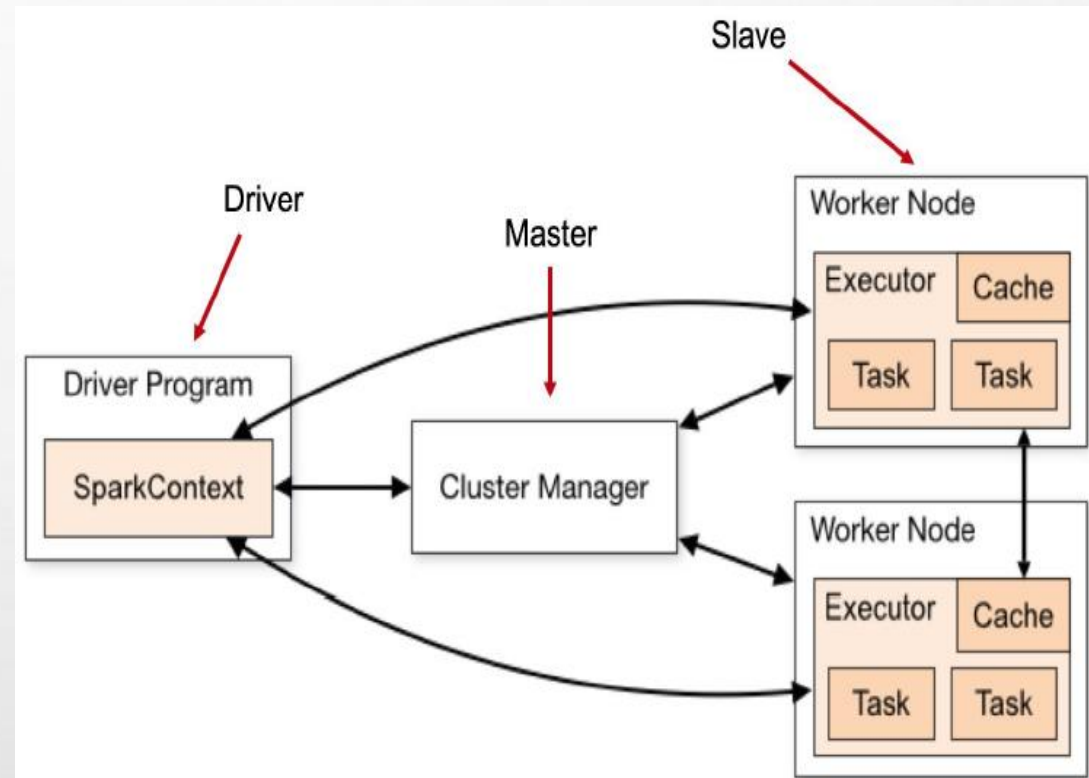
- **ECHTZEITANALYSE UND PROGNOSEN ÜBER ANGEBOT, VERHALTEN, LAGERBESTÄNDE USW.**
- **STEIGERN LOGISTIK UND SUPPLY CHAIN MANAGEMENT EFFIZIENZ BEI DER BEREITSTELLUNG, BESCHAFFUNG, TRACKING VON PRODUKTEN**
- **AGGREGATION VON INFORMATIONEN, UM BESSERE ENTSCHEIDUNGEN IM UNTERNEHMEN VORHERZUSAGEN.**
- **SCHNELL STEIGENDE VERBRAUCHSDATEN (FÜR BEISPIEL MOBILE) ERHALTEN DURCH VIELFALT DER QUELLEN**



Supply Chain Management managing complex and dynamic supply demand Network(cf. Wieland \ Wallemberg,2011)[7]

APACHE SPARK ALS BESTE LÖSUNG

- **WIE SCHON ERWÄHNT LÄUFT SPARK 100 MAL SCHNELLER ALS HADOOP. [8]**
- **ES VERFÜGT ÜBER EINE ERWEITERTE DAG EXECUTION-ENGINE**
- **UNTERSTÜTZUNG ZYKLISCHEN DATENFLUSS UND IN-MEMORY-COMPUTING.**
- **SPARKCONTEXT WIRD DEFINIERT UND WIRD MIT DEN RDD (RESILIENT DISTRIBUTED DATASET) INITIALISIERT**



Architecture of Spark Stream[9]

SPARKCONTEXT

- **SPARKCONTEXT BEKOMMT DIE RDD UND STELLT DIE SALVE ZUR VERFÜGUNG**
- **DIE AUFGABEN WERDEN AN SLAVE VERTEILT**
- **DER DATENAUSTAUSCH WIRD DURCH RDD GESICHERT**
- **SPARK WIRD NICHT DIREKT AUF DEM DISK SCHREIBEN**

```
1  
2 from pyspark import SparkContext, SparkConf  
3 conf = SparkConf().setAppName("crime-treatment").setMaster("local")  
4 sc = SparkContext(conf=conf)  
5
```

Initialisierung der Sparkcontext Beispiel

```
1  
2 data = sc.textFile("file:/app/crimes_chicago_2012_2015.csv")  
3
```

RDD Initialisierung Beispiel

ZIEL DES FORSCHUNGSBEREICHS

- **EINE BIG DATA DATEN UMGEBUNG EINRICHTEN**
- **ANWENDUNGSSPEZIFISCHES GESCHÄFTSMODELL DEFINIEREN**
- **PROAKTIVE LÖSUNGEN FÜR DAS SPEICHERN UND ZUGRIFF AUF DATEN IM BIG DATA BEREICH DEFINIEREN**

MEHR WERTE DURCH BIG DATA



FAZIT

- **ES WAR EINFACH, DAS THEMA ZU DEFINIEREN AUFGRUND DES VORHERIGEN BACHELORSTUDIUMS**
- **DAS EINSTEIGEN IM BIG DATA ENGINEERING BLEIBT NOCH EINE HERAUSFORDERUNG**

LITERATURVERZEICHNIS

- [1] [HTTPS://DE.WIKIPEDIA.ORG/WIKI/SEMISTRUKTURIERTE_DATEN](https://de.wikipedia.org/wiki/Semistrukturierte_Daten)
- [2] [HTTP://DE.MARKLOGIC.COM/BLOG/MULTI-MODEL-DATENBANKEN/](http://de.marklogic.com/blog/multi-model-datenbanken/)
- [3] KAISLER, S RÜSTUNG, F ESPINOSA, J, A UND GELD, W, 2 013, GROßE DATEN: HERAUSFORDERUNGEN UND FORTSCHRITTE VORWÄRTSBEWEGUNG, SYSTEMWISSENSCHAFTEN (HICSS), 2013, 46. HAWAII INTERNATIONALE KONFERENZ AM 7-10 JAN 2013, 995 - 100
- [4] ABDELKARIM BEN AYED, MOHAMED BEN HALIMA, ADEL M. ALIMI 2015. BIG DATA ANALYTICS FOR LOGISTICS AND TRANSPORTATION.2015 4TH IEEE INTERNATIONAL CONFERENCE ON AVANCED LOGISTICS AND TRANSPORT(ICALT), S.311.
- [5] K. WEDGWOOD AND R. HOWARD, "BIG DATA AND ANALYTICS IN TRAVEL AND TRANSPORTATION", IBM BIG DATA AND ANALYTICS WHITE PAPER, NOVEMBER 2014 .
- [6] NAZIM H. MADHAVJI, ET AL.2015. BIG PICTURE OF BIG DATA SOFTWARE ENGINEEING.2015 1ST IEEE INTERNATIONAL WORKSHOP ON BIG DATA SOFTWARE ENGINEERING, S.11.

LITERATURVERZEICHNIS

- [7] ANDREAS WIELAND, CARL MARCUS WALLENBURG (2011): SUPPLY-CHAIN-MANAGEMENT IN STÜRMISCHEN ZEITEN. BERLIN
- [8] MR. HARJEET SINGH JAGGI, MR. SUNNY S. KADAM*, “ INTEGRATION OF SPARK FRAMEWORK IN SUPPLY CHAIN MANAGEMENT ”, 7TH INTERNATIONAL CONFERENCE ON COMMUNICATION, COMPUTING AND VIRTUALIZATION (ICCCV 2016), S.1018
- [9] [HTTP://SPARK.APACHE.ORG/DOCS/LATEST/CLUSTER-OVERVIEW.HTML](http://SPARK.APACHE.ORG/DOCS/LATEST/CLUSTER-OVERVIEW.HTML)(09:03:2017)

DANKESCHÖN FÜRS ZUHÖREN