



Mit Beiträgen von  
Branchenführern  
wie Otto, Siemens,  
Mercedes u.v.m.  
- Auf über 170 Seiten -

# WHITEPAPER

## „DATA SCIENTIST – DIE KARRIERE DER ZUKUNFT“

### WIE DIE ANALYSE VON DATEN DIE WIRTSCHAFT PRÄGT



## Inhaltsverzeichnis

• Absatzprognosen im Versandhandel – Otto Group	3
• Big Data verändert den Einzelhandel – Sutterlüty	20
• Logistik und Passagierbewegungen – Fraport	38
• Big Data im Gesundheitssektor – Charité	64
• Big Data in der Finanzwirtschaft – Kreditech	87
• Sensordatenanalyse – Siemens	115
• Echtzeitauswertung von Messdaten – Mercedes	126
• Sonstige Beispiele – Capgemini	140
• Big Data Architekturen – Microsoft	159

## Data Science Best Practices

Unternehmen haben in den vergangenen Jahren eine enorme Menge an Daten (Big Data) angesammelt. Daraus zukunftsentscheidende Schlüsse zu ziehen ist Gold wert. Doch die klassischen Informatiker und Statistiker sind mit Anfragen der Fachabteilung überfordert, beispielsweise mit dieser: „Nimm diese 300TB und hole mir 10 Kunden daraus, mit denen wir in den kommenden Jahren den meisten Umsatz machen.“ Analysen von Big Data vermitteln richtungsweisende und teilweise revolutionäre Erkenntnisse für Produktoptimierung, Kreierung neuer Produkte, Markenwahrnehmung oder Financial Forecasting.

Nur der Data Scientist ist in der Lage, Big Data effektiv zu verarbeiten. Er übernimmt das Organisieren der Daten und das Bauen analytischer Modelle im Rahmen eines Projektes. Unternehmen, die von den enorm großen Datenvolumina profitieren wollen, brauchen künftig spezifisch ausgebildete Datenwissenschaftler. Laut dem IT-Analysten Gartner werden bis 2015 durch Big Data 4,4 Millionen neue Jobs entstehen.

Dieses EUROFORUM-Whitepaper mit Fallstudien aus IT (**Microsoft**) und Logistik (**Fraport**) sowie Automobilindustrie (**Mercedes**), Handel (**OTTO**), dem Gesundheitssektor (**Charité Berlin**) und weiteren Branchen beweist: Data Science ist DAS branchenübergreifende Zukunfts-Thema der Wirtschaft.

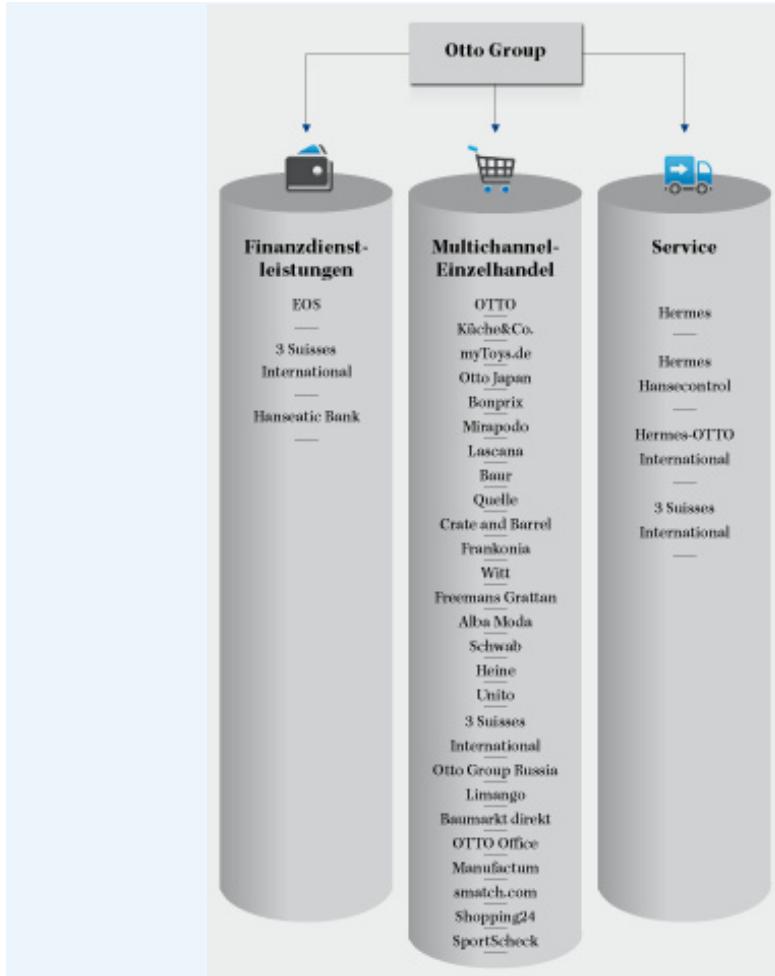
# **BIG Data & Predictive Analytics – der Nutzen von Daten für präzise Prognosen und Entscheidungen in der Zukunft**

**BIG Data Europe in Zürich, 28.08.2012**

**Otto (GmbH & Co. KG)  
Michael Sinn**



# Das vielfältige Portfolio der Otto Group



Die Geschäftstätigkeit der Otto Group erstreckt sich auf die drei Segmente:

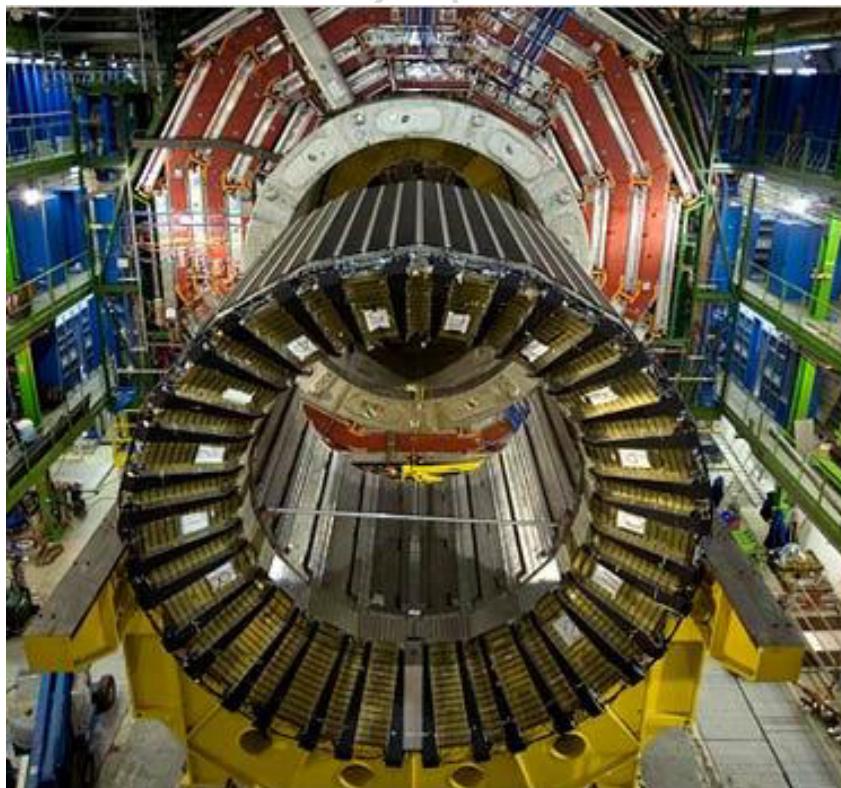
- **Multichannel-Einzelhandel –** mit Kataloggeschäft, E-Commerce und stationärem Einzelhandel.
- **Finanzdienstleistungen –** mit handelsnahen Finanzdienstleistungen entlang der Wertschöpfungskette.
- **Service –** mit dem kompletten Portfolio an Handelsdienstleistungen entlang der Wertschöpfungskette sowie Reisedienstleistungen.

# Mit vielfältigen Marken zum Erfolg

Multichannel-Einzelhandel

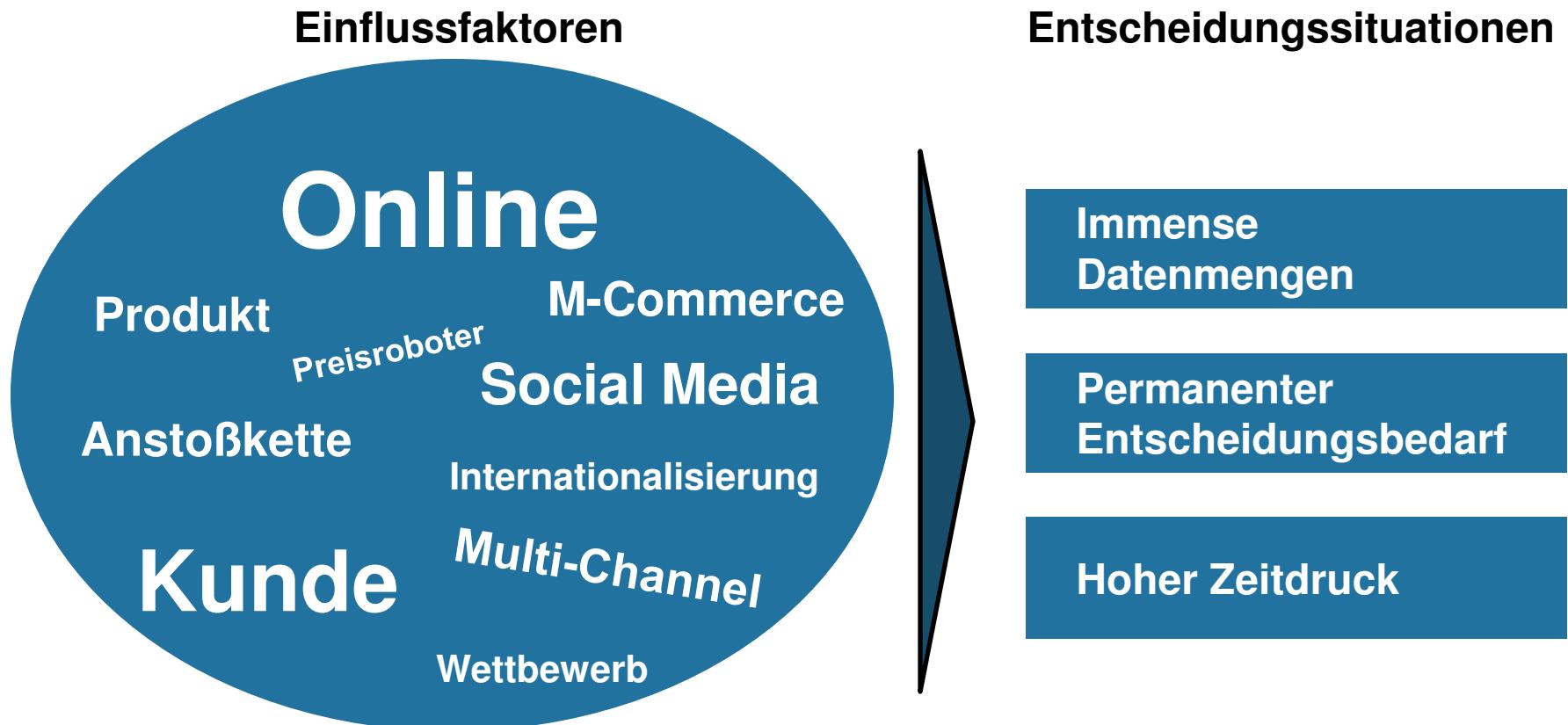


# Was hat ein Algorithmus aus der Teilchenphysik mit Versandhandel zu tun?



A screenshot of the OTTO website homepage. The top features the OTTO logo in red. Below it is a search bar with the placeholder "Suchbegriff / Artikelnummer eingeben" and a "SUCHEN" button. A banner for "TOP-MARKEN bis zu 50% REDUZIERT" (Top Brands up to 50% reduced) is prominently displayed, listing brands like ESPRIT, PUMA, REMINGTON, HIS, ASUS, and PHILIPS. Another banner offers "Last-Minute-Shopping!" with a 24-hour delivery guarantee for 1€. The page also includes sections for "Sommer kommt von OTTO", "Artikelnummern direkt eingeben", and "Mein OTTO". A "Interessante Artikel für Sie" section shows images of women in bikinis and household items. At the bottom, there are links to various catalogues and promotional offers like "extra % outlet ohntrend" and "trend".

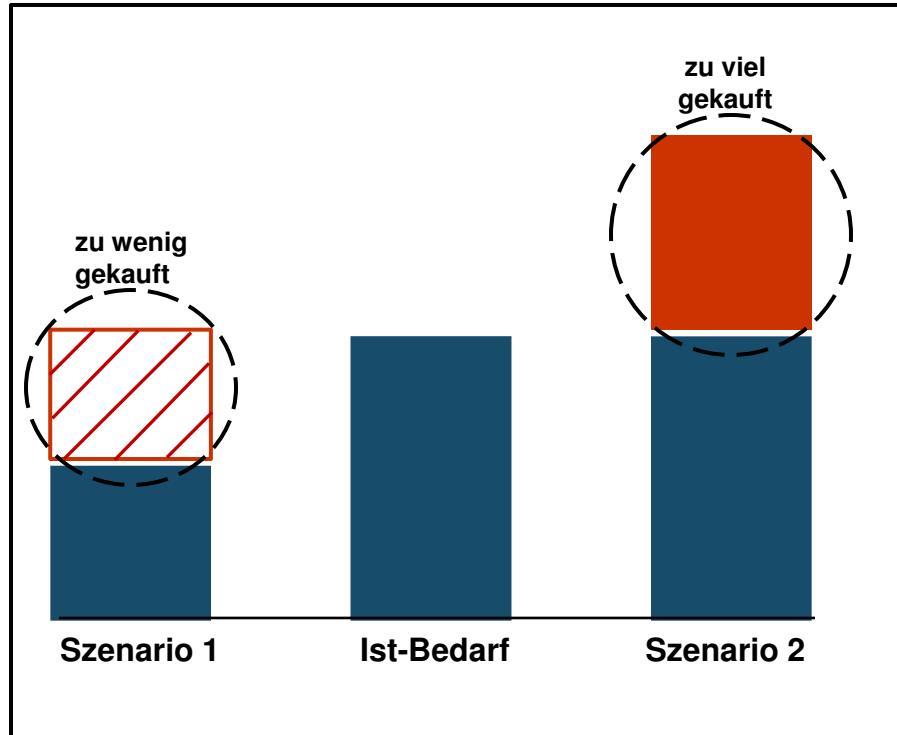
# Große Datenmengen und eine Vielzahl von Einflussfaktoren prägen heutige Entscheidungssituationen



## Determinanten für Unternehmenserfolg

# Herkömmliche Verfahren werden den gestiegenen Anforderungen an die Artikelprognose nicht mehr gerecht

## Auswirkung Dispositionentscheidungen



### Szenario 1: zu wenig gekauft

- Umsatzausfall
- Erhöhte Beschaffungskosten
- Kundenverärgerung
- Erhöhte Kontaktkosten Callcenter

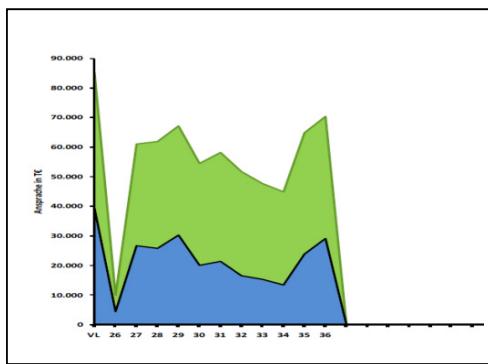
### Szenario 2: zu viel gekauft

- Erhöhte Lagerkosten
- Kapitalbindung
- Verwertungsverluste

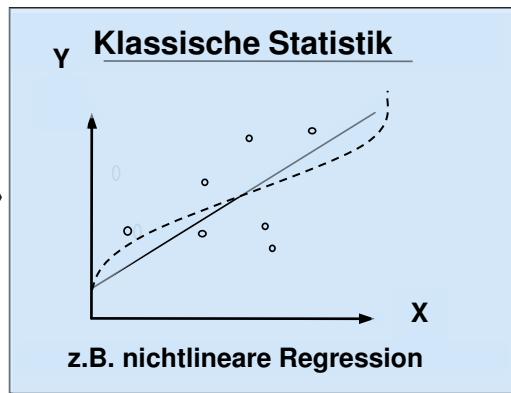
→ Es besteht eine permanente Herausforderung, die richtigen Mengen zu identifizieren

# Eine maßgebliche Verbesserung ist nur durch grundsätzlich neue Methoden zu erreichen

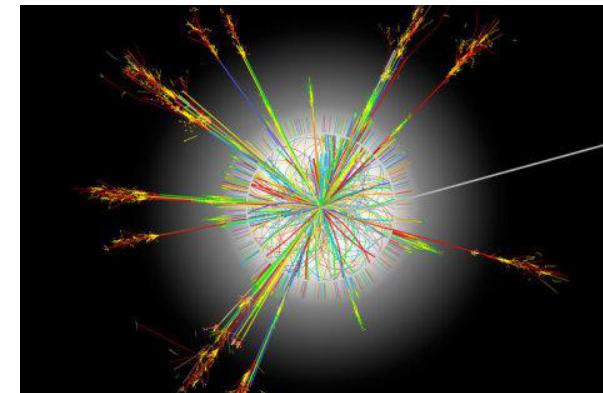
Trendmethode / Dreisatz



Klassische Statistik



Neuer Ansatz?



Forschungsauftrag an ein Experten-Team, ein Prognoseverfahren zur Abbildung der heutigen und zukünftigen Komplexität des Geschäftsmodells zu entwickeln

# Nach Evaluierung von 13 Prognosetools international ausgerichteter Softwareanbieter hat sich die NeuroBayes® Technologie von **blueyonder** durchgesetzt

Forward looking. Forward thinking.

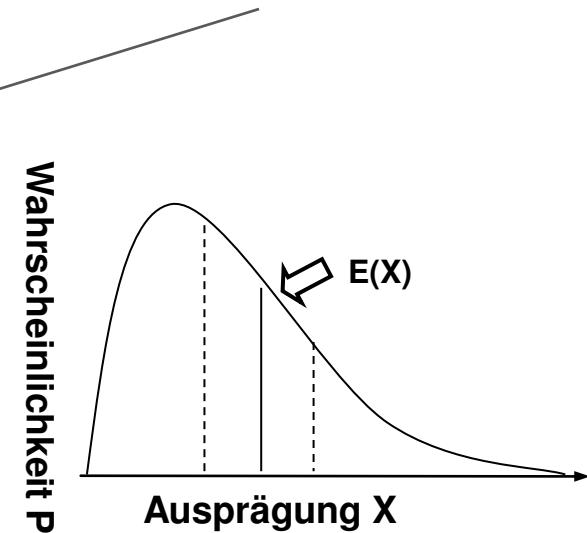
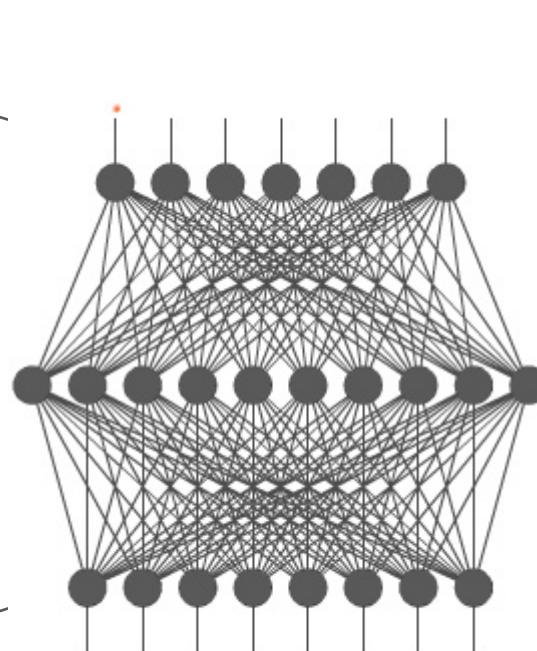
200 Inputvariablen

Aktuelle Daten

Style  
Farbe  
Marke  
Preis  
Onlineplatzierung  
Seitenanteil  
Auflage  
Response  
Bisheriger Saisonabsatz  
.....

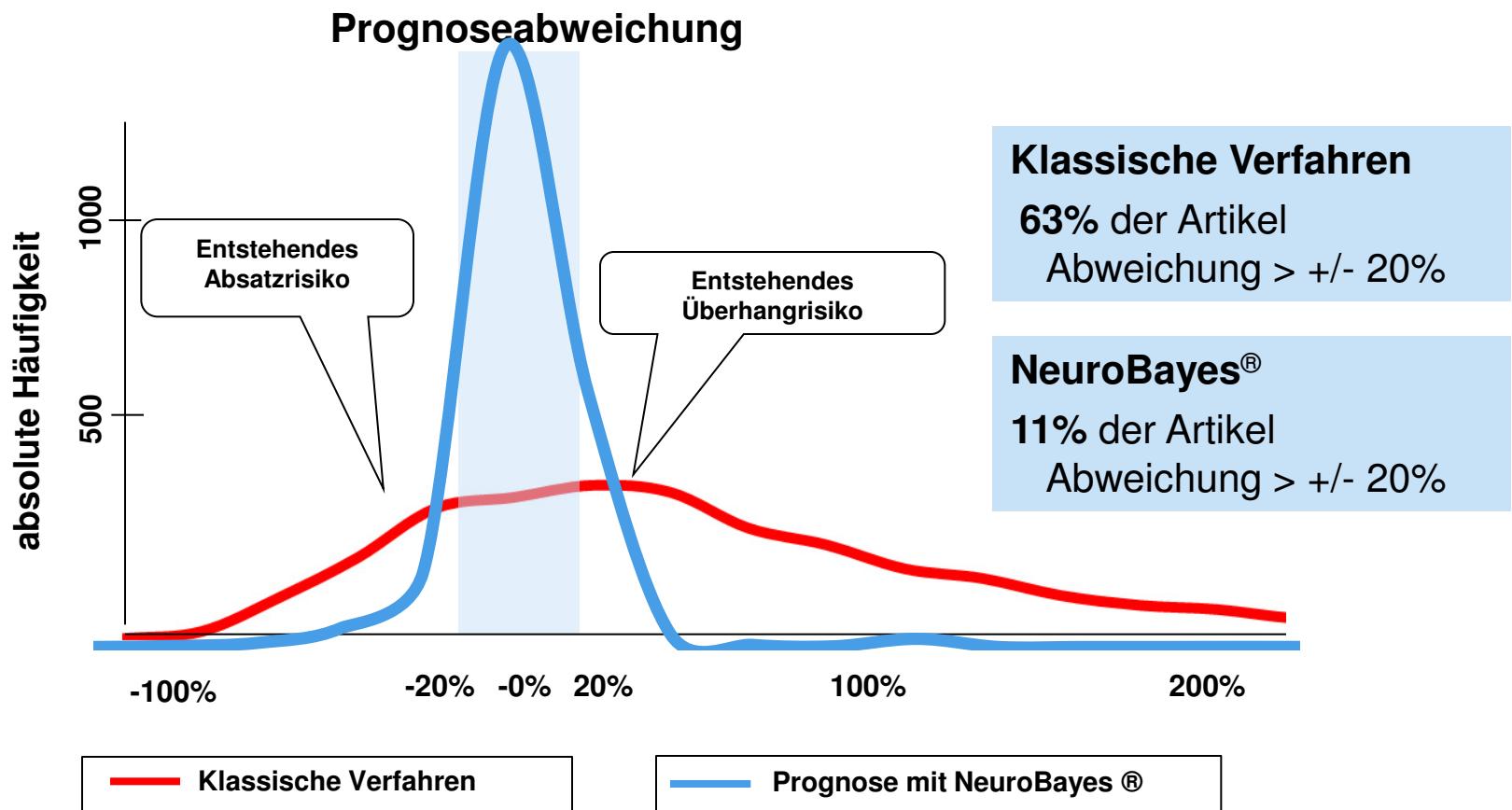
Historische Daten

16 Saisons



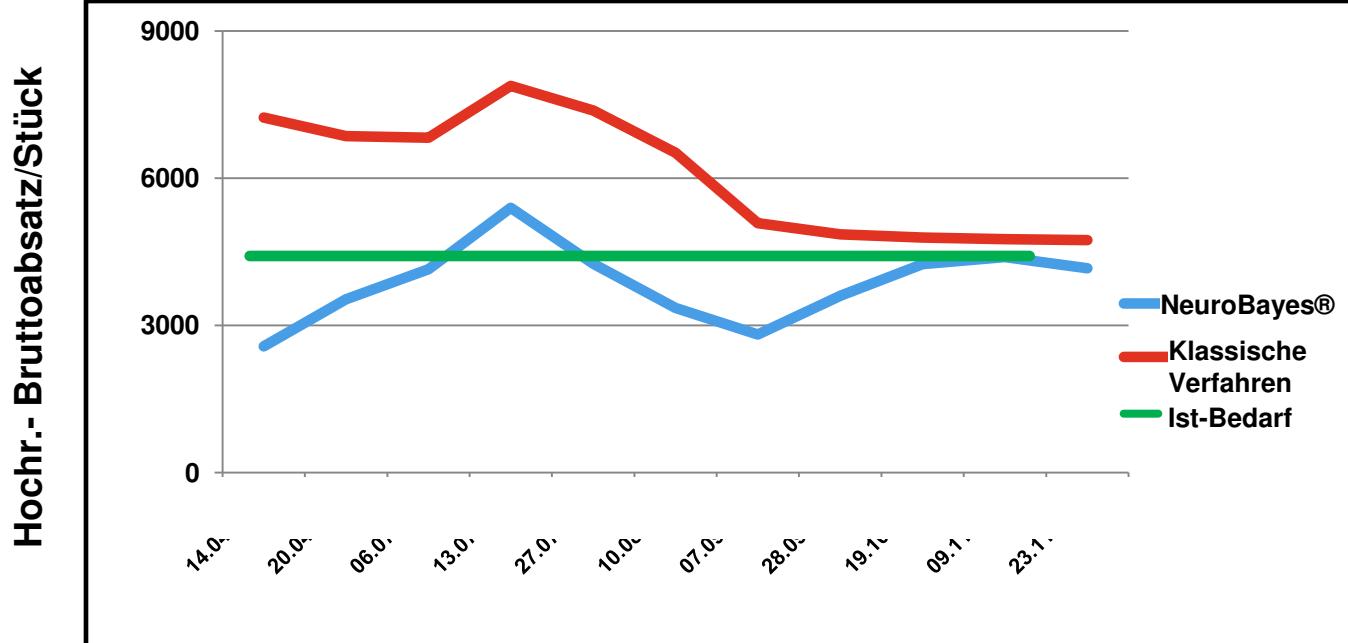
Prognose je Artikel, Farbe, Größe, Promotion und Woche jeweils für Ansprache, Brutto-, Nettoabsatz und Retouren

# Eine Gegenüberstellung der Prognoseverfahren zeigt die deutliche Verbesserung durch NeuroBayes®



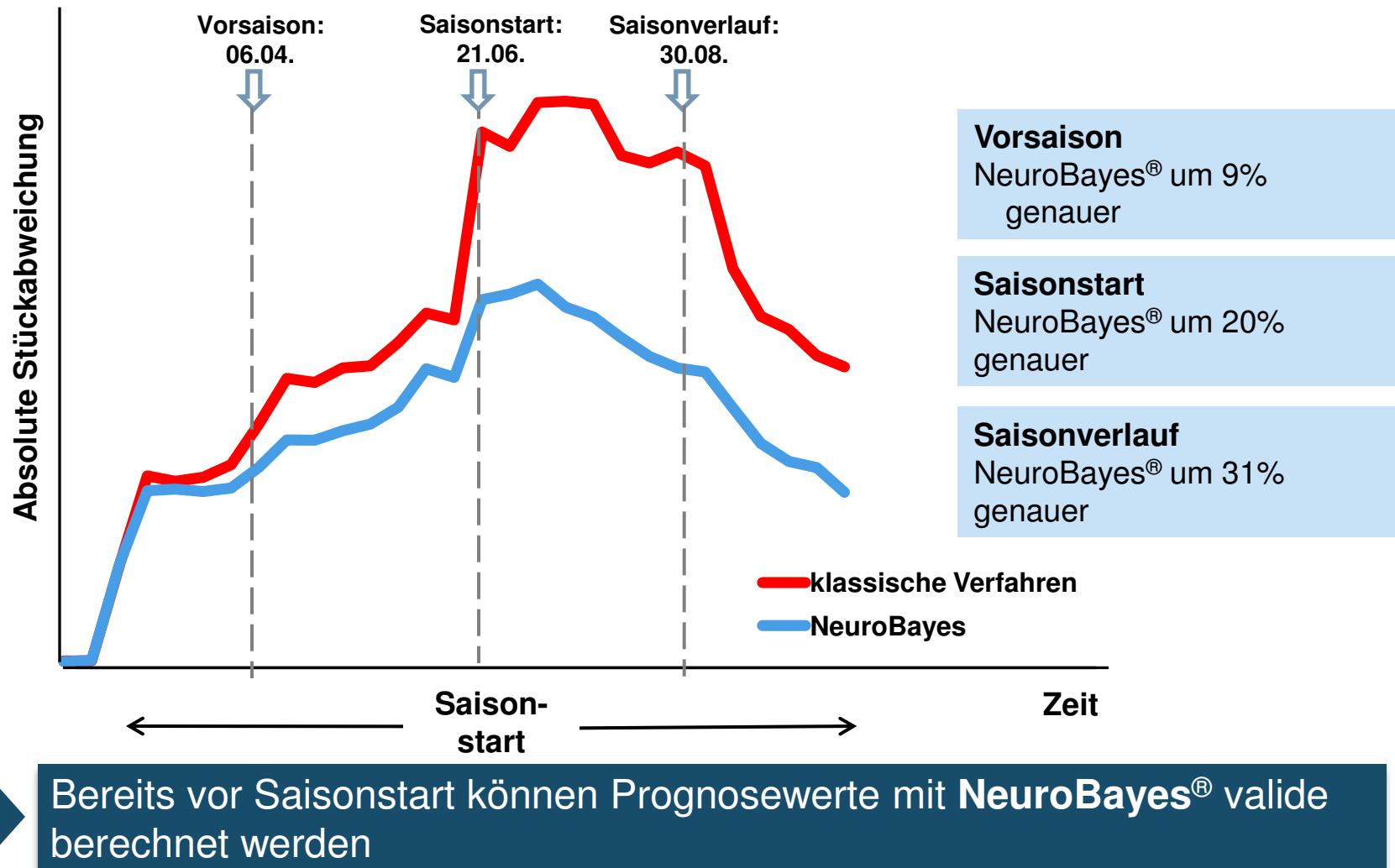
Ergebnis NeuroBayes®: Deutlich positive Wirkung auf Lieferbereitschaft und Restevolumen

# Die Artikelprognosen bilden die Basis für Kaufentscheidungen und Bestandsmanagement



**NeuroBayes®: Deutlich bessere Prognosen für sämtliche Artikelpositionen, d.h. für Farbe, Größe und Angebotsträger**

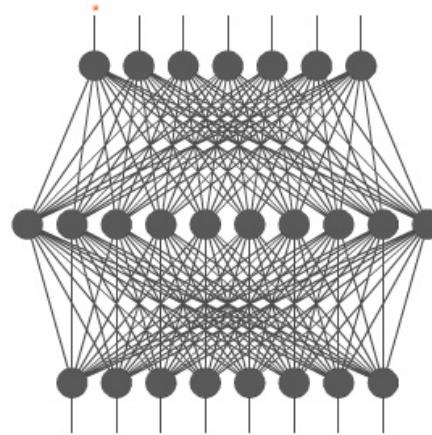
# Die größten Optimierungspotenziale können bei Online-Prognosen gehoben werden



# NeuroBayes® ist ein lernendes System, bei dem die Prognosegüte im Zeitablauf permanent besser wird

## Input

- 300 Millionen Datensätze pro Woche
- 135 GB historische Daten



## Output

- > 1 Milliarde Einzelprognosen pro Jahr

► Ab Q3 2012 erfolgt die Umstellung der Lieferung von Prognosedaten von wöchentlich auf täglich

# Die Weiterentwicklung des Prognosetools zu einem Dispotool basiert auf einem Warenmanagementsystem



Bedarfs-  
prognose

- Artikel mit Handlungsbedarf
- Dispositionsvorschlag

- Ziel:  
„automatisierte“  
Disposition

Dispositionsvorschlägen liegen Vorgaben und Kostenfunktionen für Lieferbereitschaft und Restevolumen zugrunde

# Otto gewinnt den „Retail Technology Award Europe“ in der Kategorie „Best Enterprise Solutions“



# Treffsichere Prognosen unterstützen verantwortungsvolles und nachhaltiges Wirtschaften

## Bedarfsgerechte Produktion

- Die punktgenaue Produktion von Waren reduziert den Ressourcenverbrauch.
- Die Vermeidung von Überproduktion verringert die Entsorgungsmengen deutlich.

## Zielgerichtete Verteilung

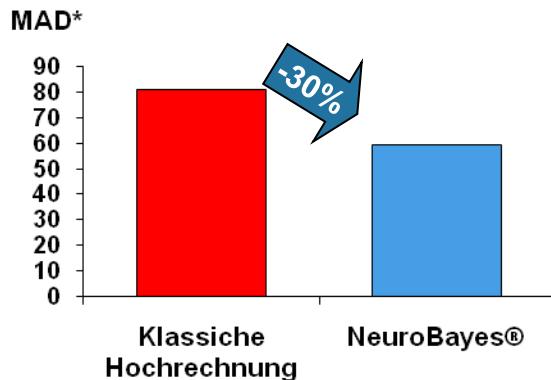
- Durch die Vermeidung/Optimierung von Transporten wird ein relevanter Beitrag zur CO2-Reduzierung geleistet.



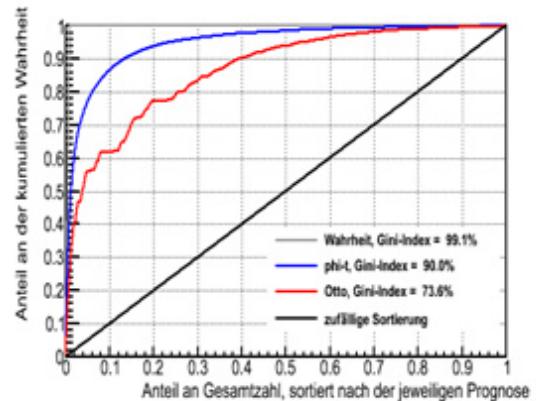
**Predictive Analytics als ein Schlüsselfaktor für Corporate Social Responsibility**

# Basierend auf der NeuroBayes®-Technologie lassen sich Grundsatzfragestellungen des Distanzhandels besser lösen

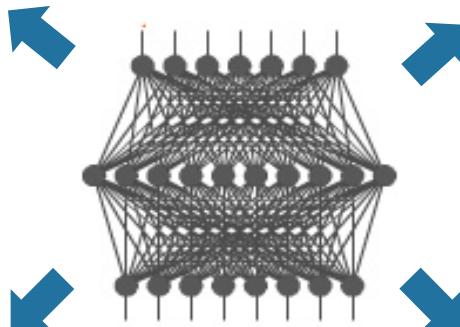
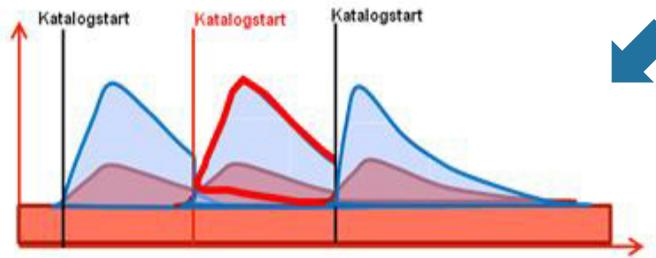
## Artikelbedarf



## Kundenbonität



## Werbemittelausstattung



## Recommendation Engine



\*Mittlere absolute Prognose-Abweichung in Stück auf Artikel-Promotion-Größen-Ebene; MAD

# Vielen Dank für Ihre Aufmerksamkeit!

Michael Sinn  
Direktor Angebots- und Category Management Support, Otto (GmbH & Co. KG)  
Email: [michael.sinn@otto.de](mailto:michael.sinn@otto.de)  
Tel: (040) 6461-6570

# Next Generation Retail

Wie Big Data den Einzelhandel verändert

**Big Data 27. August 2013**

**Mag. Philipp Giselbrecht**

Leiter Marketing, Sutterlüty Handels GmbH

**Dr. Thorben Keller**

Gründer & CEO, Cosibon AG



# Sutterlüty in Vorarlberg



Anzahl Sutterlüty Geschäfte: 22  
Anzahl Sutterlüty Mitarbeiter: 700  
Umsatz Sutterlüty Gruppe 2012: 109 Mio. Euro









# Herausforderungen im CRM eines stationären Lebensmittelhändlers

- Anzahl digitaler Werbekanäle explodiert
- Marketingbudget bleibt unverändert
- Sinn einer physischen Kundenkarte
- Smartphone-App als Kundenbindungsinstrument
- Steigerung der Bedarfsdeckungsquote
- Wie bewerten die Kunden unsere Produkte, welche Empfehlungen sprechen Sie aus
- Wie schaffe ich eine Interaktion mit dem Kunden
- Einsparungen in der 1:1 Kommunikation
- Echter Mehrwert eines Apps für Kunden
- Kann ein App die Welt ein wenig „grüner“ zu machen
- Messung der Werbewirksamkeit
- Erhöhung von Markenkontaktpunkten

# **Bedürfnisse eines Kunden im lebensmitteleinzelhandel**

- Herkunft von Produkten
- Kassenzettel als Kontrollinstrument
- Bequeme Einkaufslisten-Erstellung
- Persönlicher Beitrag zum Umweltschutz
- Gesunde Ernährung
- Übersicht der Lebensmittel-Ausgaben
- Information über Angebote
- Produktempfehlungen anderer Kunden

# Die Sutterlüty App



# Kundenkarte als Mittel zur Informationsgewinnung



**1 Messbarer Kundenkontakt pro Einkauf**  
Ort und Zeit des Einkaufs  
Gekaufte Produkte

# Kundenkarte als Mittel zur Informationsgewinnung



Aber was passiert in der Zwischenzeit?

# Der Kunde ist nicht untätig

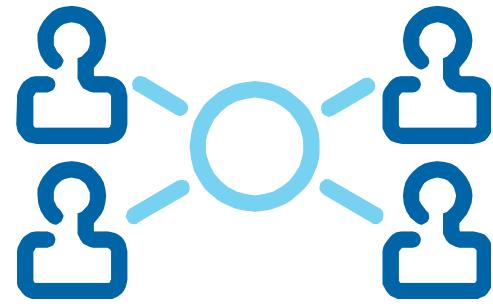
Aktuelle Angebote,  
Rabatte, Trends



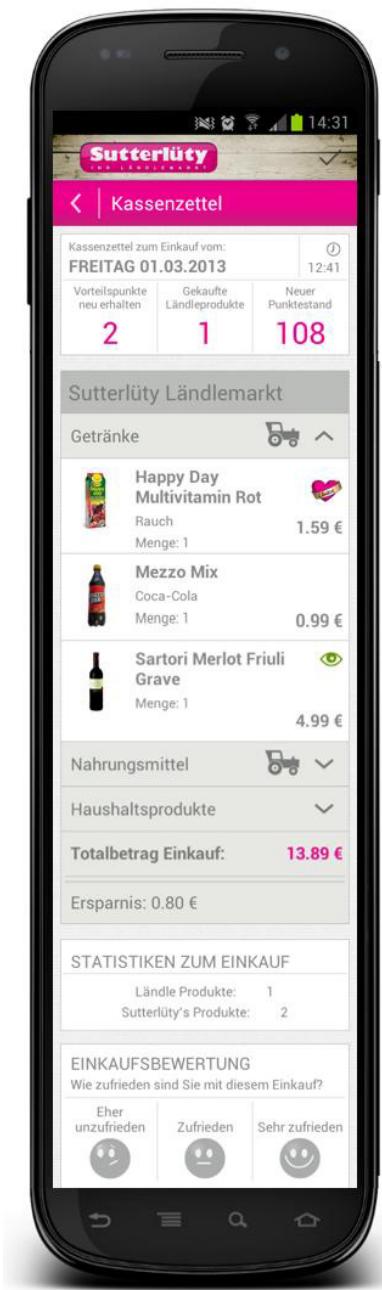
Informationen über  
Produkte, Hersteller,  
Gütesiegel



Austausch mit Freunden  
und Bekannten



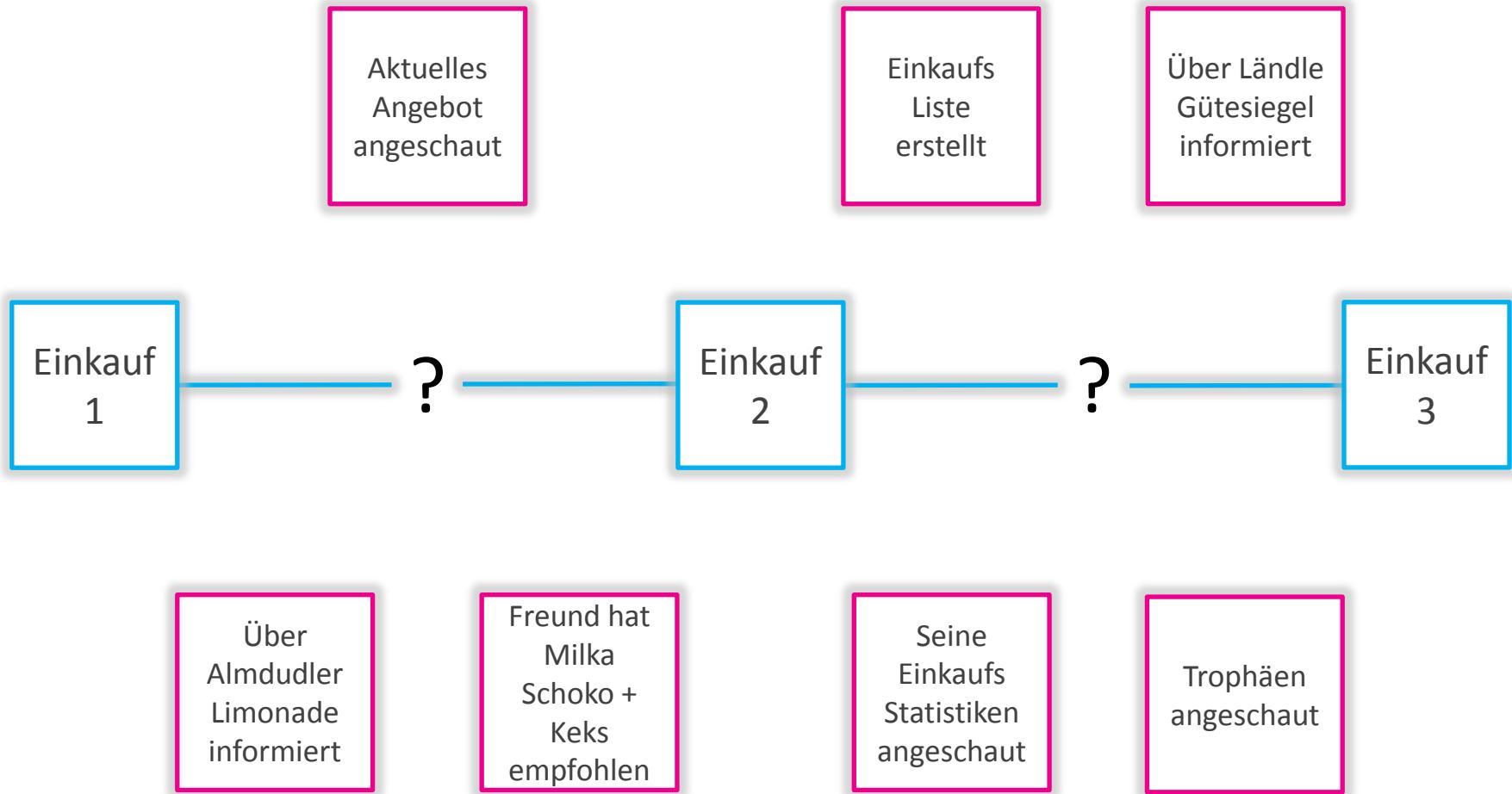
## Unternehmen haben jedoch keinen Einblick



# Beispiel



# Beispiel



# Beispiel



Aktuelles Angebot angeschaut



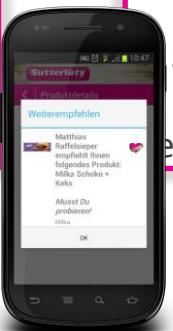
Einkaufs Liste erstellt



Über Ländle Gütesiegel informiert



Über Almdudler Limonade informiert



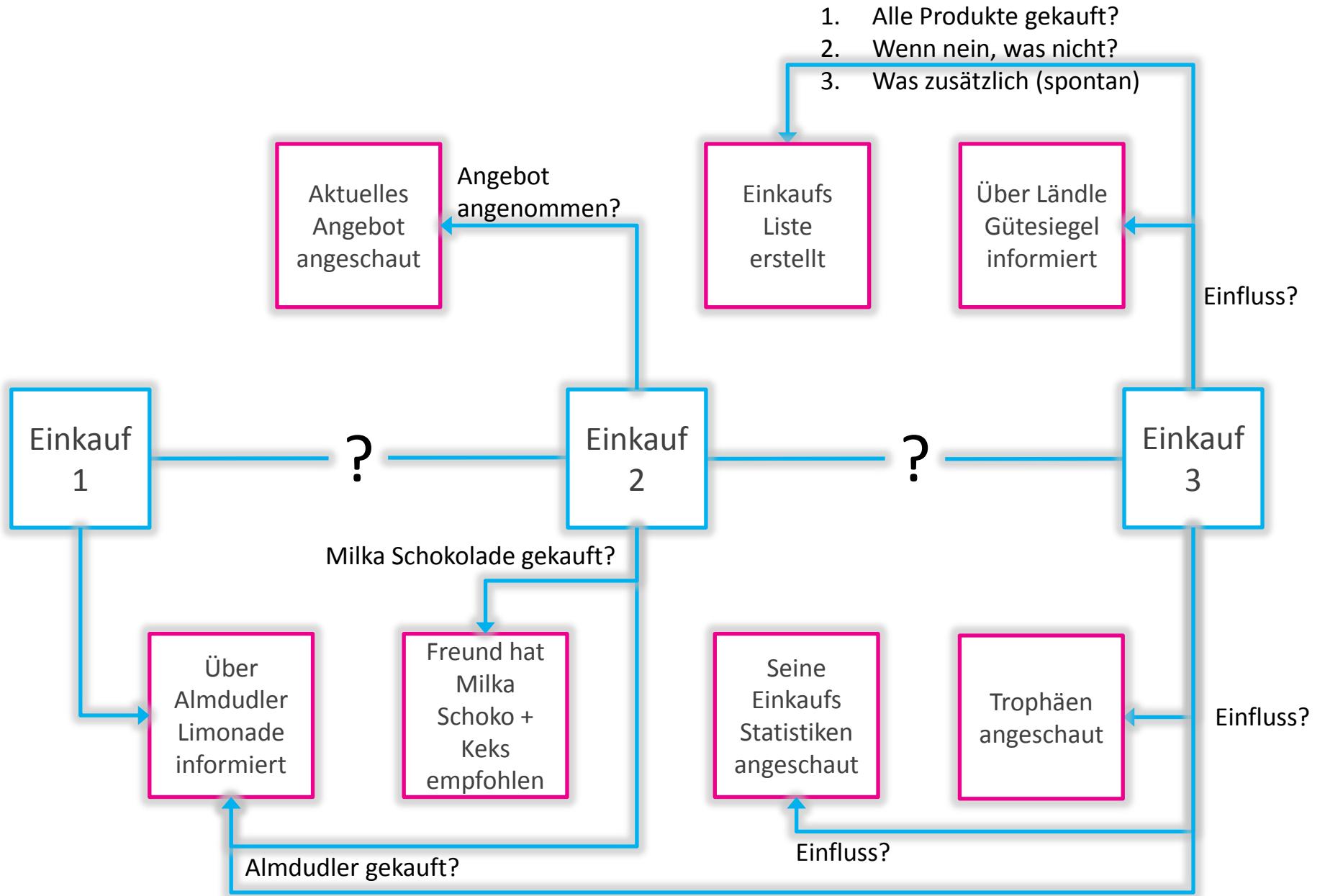
Freund hat Milka Schoko + Keks empfohlen



Seine Einkaufs Statistiken angeschaut



Trophäen angeschaut



# Big Data mit Cosibon

## (Kleine Auswahl)

1. Welchen Einfluss haben aktuelle Angebote auf welche Kundensegmente?  
(In der App sogar A/B Test möglich!)
2. Über welche Produkte informieren sich welche Segmente?
3. Welche Produkte sind für Kunden interessant, aber nicht im aktuellen Sortiment?
4. Einfluss von Empfehlungen aus dem sozialen Netzwerk
5. Welche Produkte schreiben sich Kunden auf die Einkaufsliste? In welcher Reihenfolge? Was kaufen sie dann, was nicht?
6. Einfluss von zusätzlichen Informationen (z.B. Gütesiegel) auf die Kaufentscheidung?
7. Konsumsteuerung durch Gamification möglich? Wenn ja bei welchen Segmenten?
8. Einfluss von Feedback zum Einkauf (z.B. Anteil Ländleprodukte)?

Kontinuierliche Auswertung über die Zeit ermöglicht Modell um stetig höherer Nutzerakzeptanz und immer bessere Kaufberatung/Konsumsteuerung zu ermöglichen.

Bestes Modell für «Erziehung» von Kunden im Sinne des Unternehmens ist ableitbar.  
Bei Sutterlüty zum Beispiel: Regionale und gesunder Ernährung.

# Vielen DANK!

## Testzugangsdaten App:

Store: Google Play

Kundennummer: 123456

Kennwort: BigData



# *Big Data bei der Fraport AG*

*Köln, 25. November 2013  
Stefan Sabatzki*



## Inhaltsverzeichnis

- Fraport auf einen Blick
- Big Data bei Fraport
- Historie der BI-Umgebung
- Use Case Passagierflüsse



# Fraport auf einen Blick

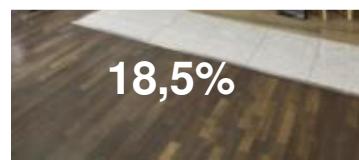
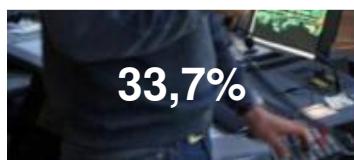
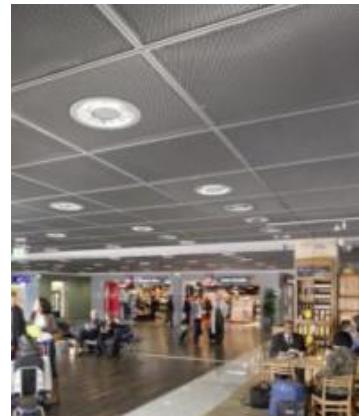
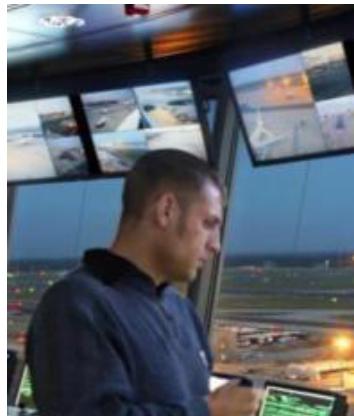
## ***Unsere Historie***

Vom Flughof Manager 1924...



...zum Manager internationaler Drehkreuze und Airports.

## *Umsatzträger sind unsere Kerngeschäftsfelder*



### Aviation

823,4 Mio. Euro

### Ground Handling

649,3 Mio. Euro

### Retail & Real Estate

452,9 Mio. Euro

### External Activities & Services

516,4 Mio. Euro

*Umsätze im Gesamtkonzern nach Segmenten 2012*

## **Frankfurt Airport – Zahlen und Fakten**

- 21 km<sup>2</sup> Fläche
- bis zu 94 Flugbewegungen/Std.
- 2 Start- und Landebahnen
- 1 Startbahn, 1 Landebahn
- 2 Fluggast-Terminals und
- 1 AIRail Terminal (ICE-Bahnhof)
- 154 Gates; 204 Positionen
- AirCargo Gleisanschluss

### **An einem Tag\*:**

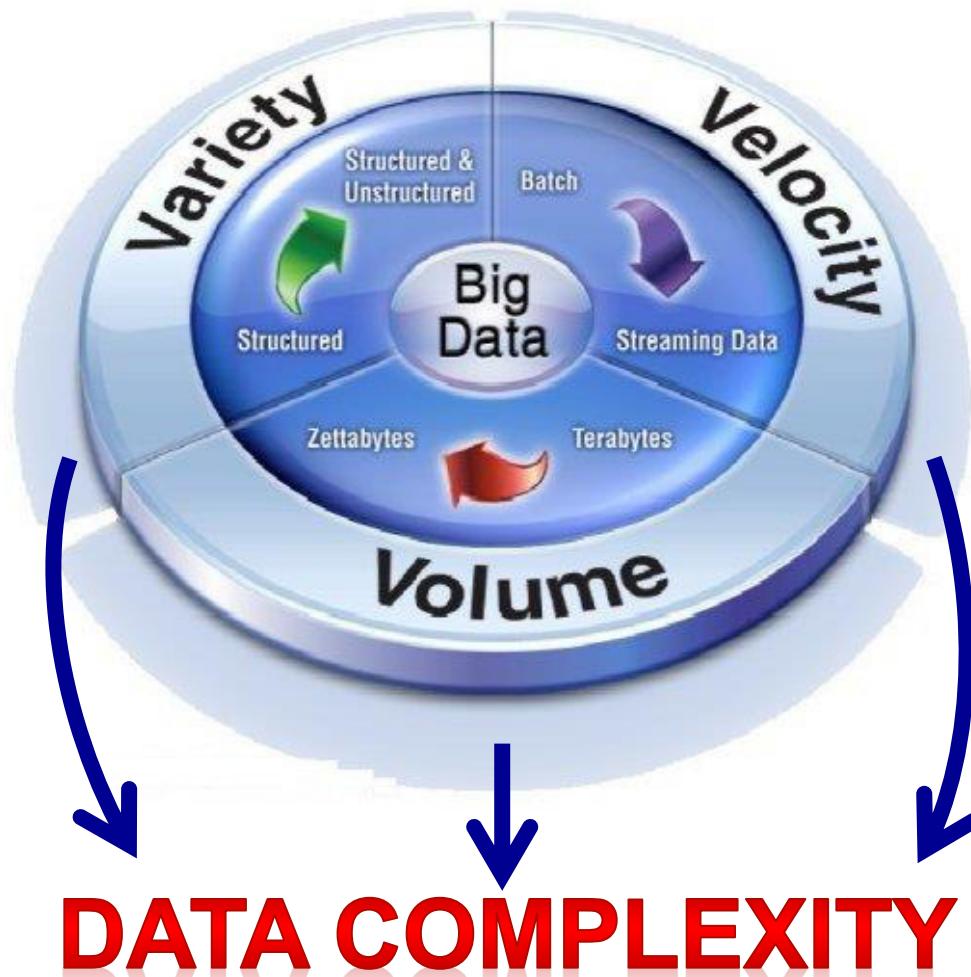
- 158.000 Passagiere
- 78.000 Stück Abfluggepäck
- 5.800 Tonnen Cargo
- 397 Züge an den Flughafen-Bahnhöfen (Fahrplan 2010)



\*durchschnittliche Angaben aus 2012

# ***Big Data bei Fraport***

# Eine Definition (von vielen)



## ***Big Data***

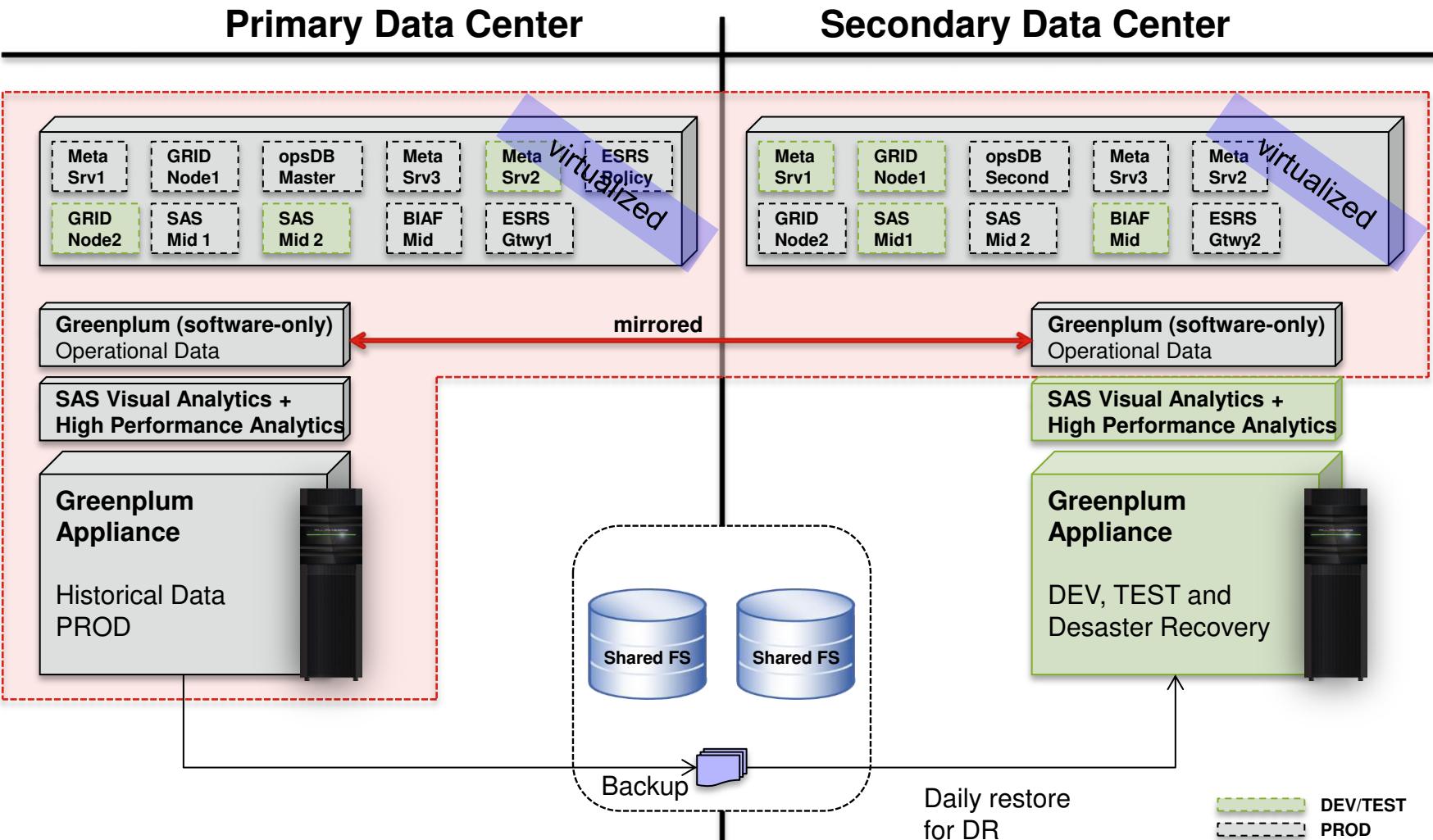
- Velocity (Geschwindigkeit)
  - Abfragegeschwindigkeit
  - Time to Market neuer Lösungen
  - Verschiedene Ladezyklen
- Variety (Vielfalt)
  - Anzahl Datenquellen
  - Unstrukturierte Daten
- Volume
  - Nicht Datengröße sondern Menge von Messages

# ***Historie der BI-Umgebung***

## Hintergrund

- Inbetriebnahme des „Business Intelligence Architecture Framework“, der BI-Umgebung für operative Daten im Jahr 2005
  - Ziel: Ganzheitliche Sicht auf Betriebszustände für das operative Management mittels
    - historischer Berichte
    - online Reports
  - Nutzerkreis erweitert sich um Disponenten, Betriebsleiter, Mitarbeiter der Betriebssteuerung
  - BIAF wird nicht mehr nur rein informativ unterstützend, sondern für die betriebliche Steuerung genutzt
  - 2013 wird BIAF in die Liste der betriebskritischen, sogenannten „Prio 1 Systeme“ aufgenommen
- **Die Verfügbarkeit des Systems wirkt sich auf die Qualität der operativen Prozesse aus. Bei Ausfall des Systems stehen wichtige Informationen zum „Systemzustand“ Flughafen Frankfurt nicht mehr zur Verfügung.**

# Systemüberblick SAS 9.4 + Greenplum



## **Hardware in Zahlen**

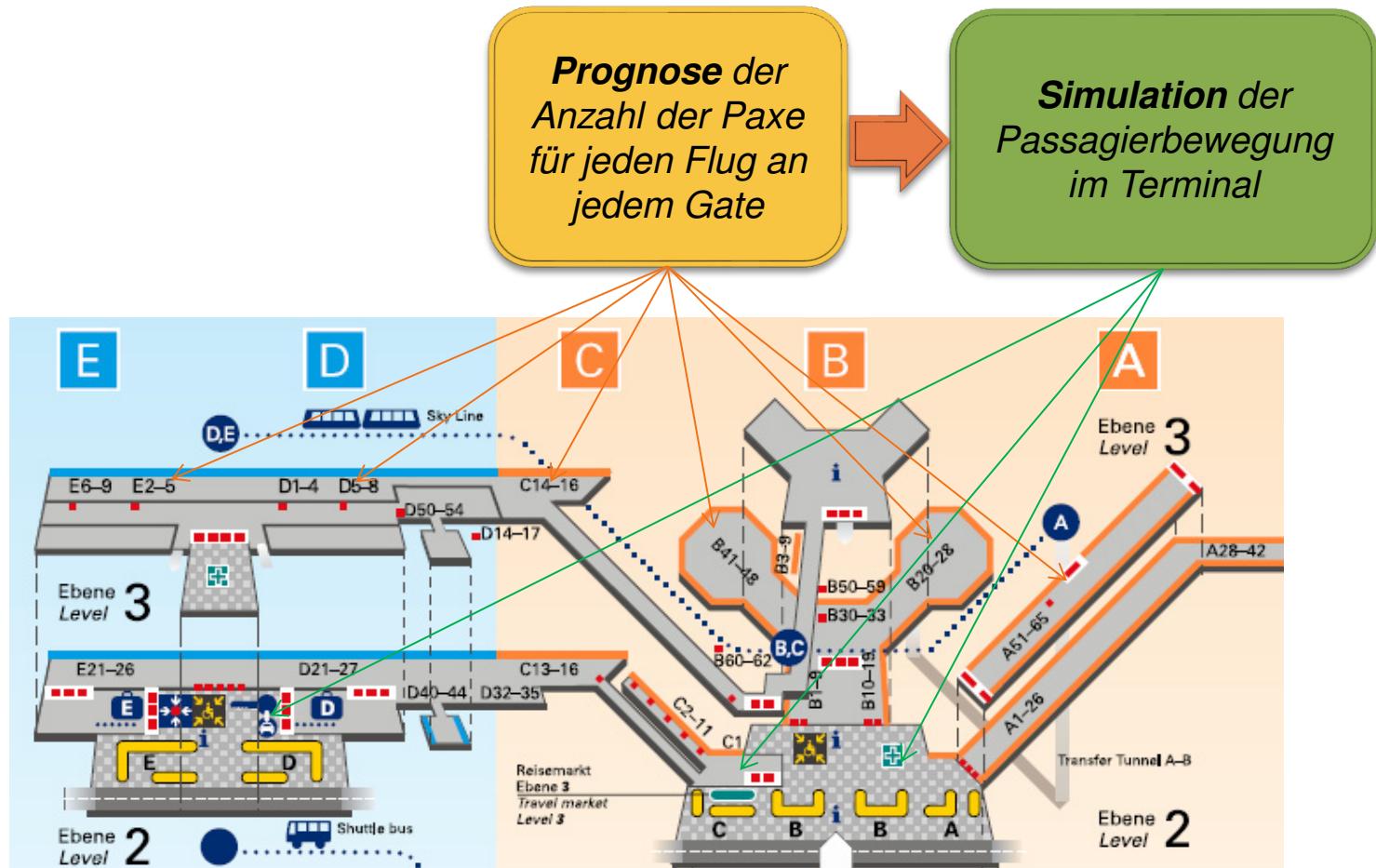
- SAS
  - 14 virtuelle Maschinen
  - 2 physische Maschinen
  - 60 Cores
  - 188 GB RAM
- Greenplum software-only
  - 2 virtuelle Maschinen
  - 2 physische Maschinen
  - 24 Cores
  - 195 GB RAM
- Greenplum Data Computing Appliance (x2)
  - 10 physische Maschinen (2 Master, 8 Segmente)
  - 96 Cores
  - 384 GB RAM

# Use Case Passagierflüsse

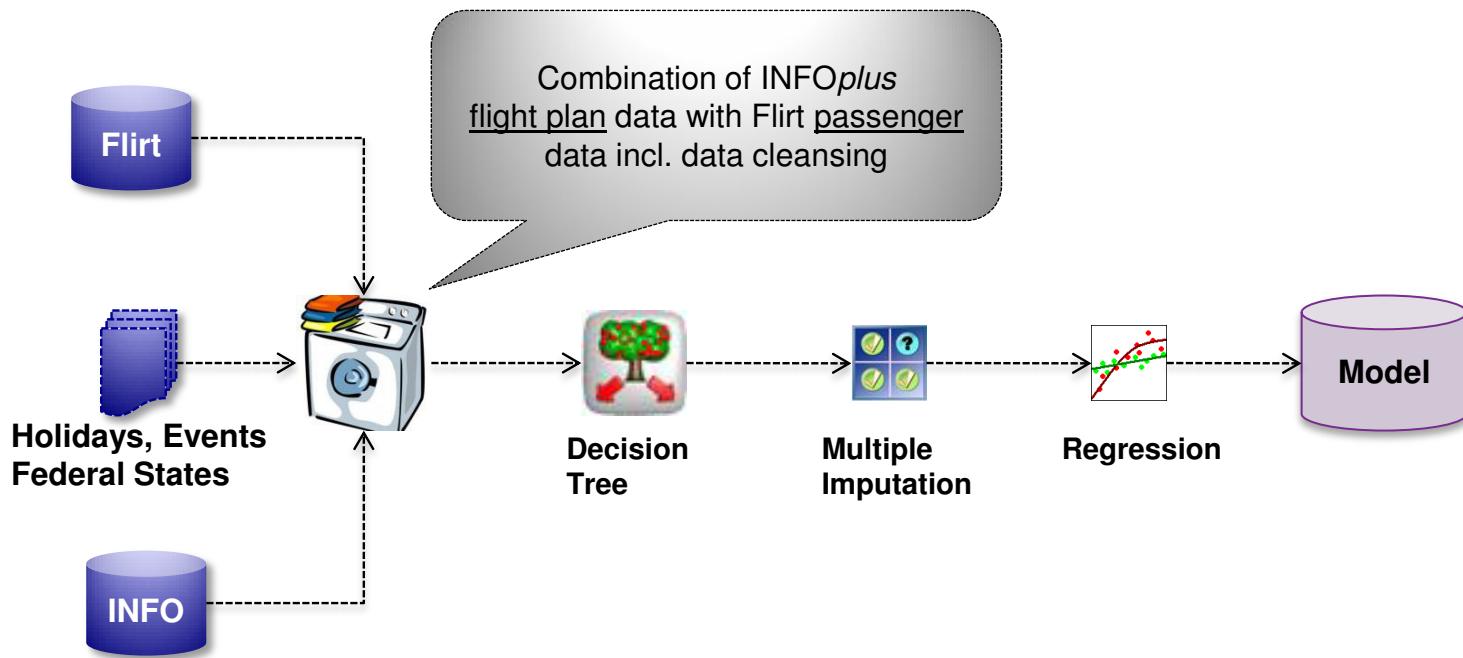
# Motivation

- Kundenseitig
  - Kostendruck der Airlines
  - Optimierung von Wartezeiten der Passagiere
- Infrastruktur
  - Platzmangel für weitere Kontrollstellen
  - Kapazitätsengpässe aufgrund von Baustellen
  - Dezentralisierung von Kontrollstellen
  - Komplexe Passagierflüsse aufgrund komplexer Topologien
- Bedarf nach höherer Servicequalität

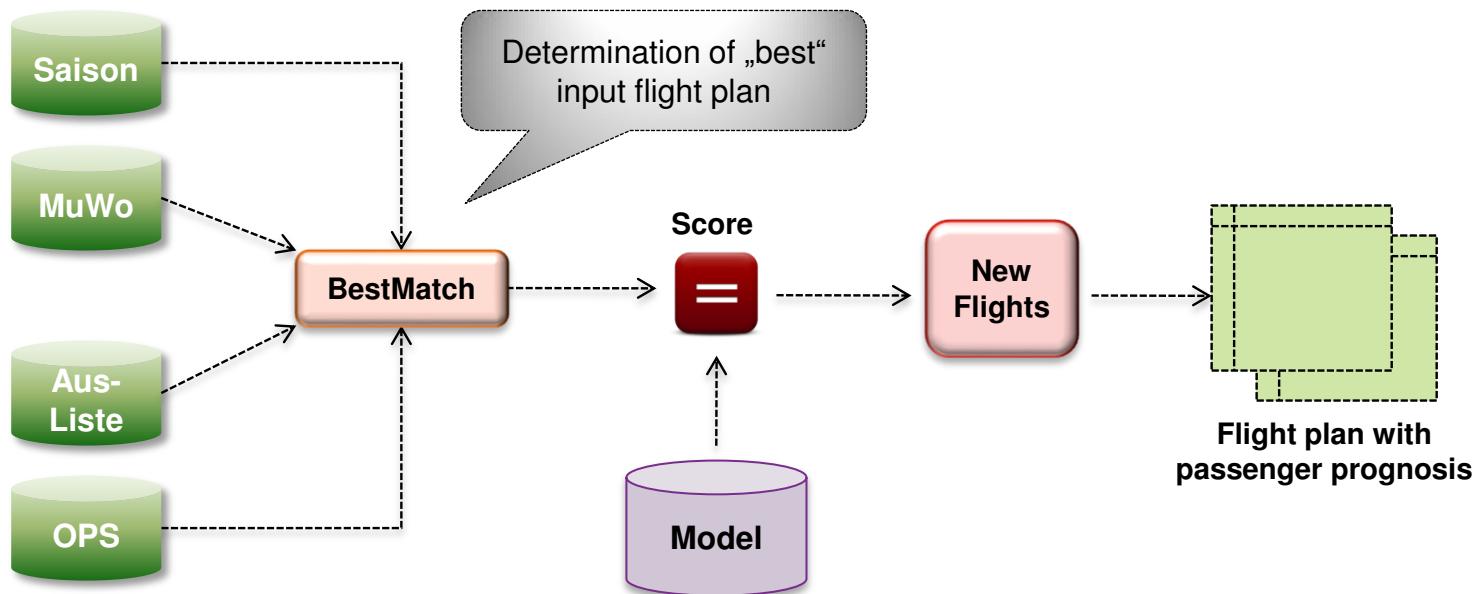
# Architektur



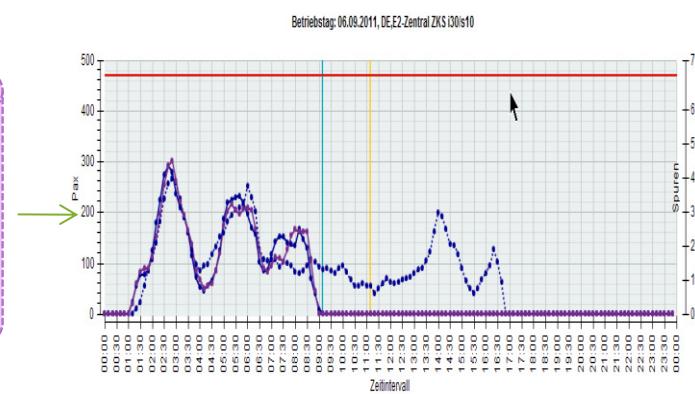
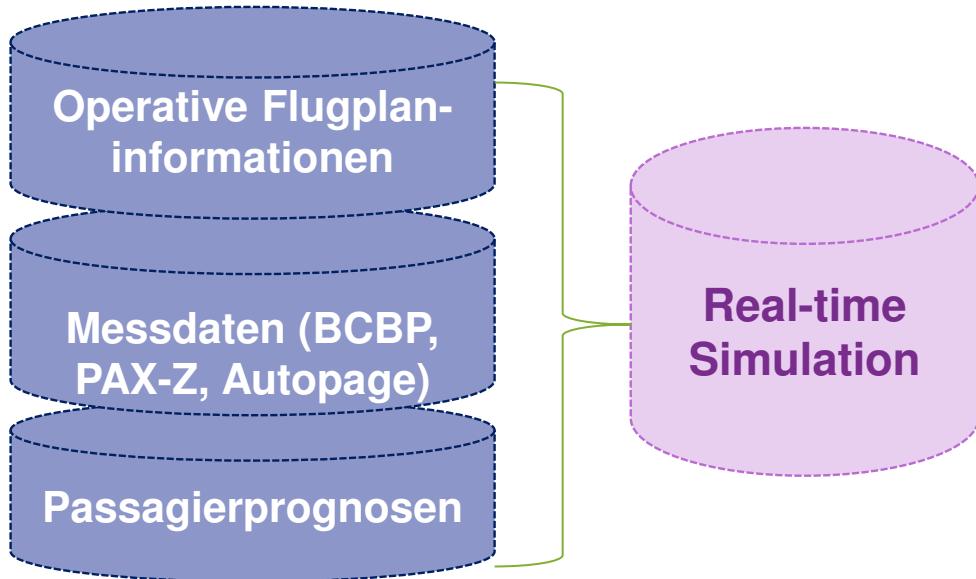
# Training des statistischen Modells



# Passagier Prognose



# ***Passagierfluss Simulation***



**Step 1:**  
**Daten Aufbereitung**

**Step 2:**  
**Simulation**

**Step 3:**  
**Operative Darstellung**

## **Passagierfluss Simulation**

- Die Simulation eines Tages erzeugt ~ 1.5 Millionen Ereignisse (Ein-, Austritte)
- Die Daten werden in der Datenbank in ~ 450.000 Datensätzen abgelegt
- 34 MB an Daten werden so pro 24h-Simulation generiert
- Bei einer Simulation, die alle 5-Minuten gestartet wird und stets den gesamten Tage simuliert, werden täglich 130 Millionen Datensätze und 9.7 GB an Daten generiert
- Stresstest im Cluster mit 24 Simulationen gleichzeitig → erfolgreich

## ***Status Quo und Vision***

- Passagierprognose pro Flugnummer
  - Ist  
Für jeden Flug der am Frankfurter Flughafen geplant ist, wird eine Prognose der Total on Board Passagiere und der Umsteiger mit der Zieldestination vorgenommen, der Prognose Lauf dauert im Moment ca. 7 Stunden und wird für unterschiedliche Prognosezeiträume angewandt.
  - Ziel  
Reduktion der Prognoseläufe in den einstelligen Minutenbereich, um die Kurzfristprognose für den operativen Einsatz in der Passagierflusssteuerung qualitativ zu verbessern.

# *Vielen Dank für Ihre Aufmerksamkeit!*



## **BACKUP**

### ***Neue Anforderungen an die Architektur***

- Höhere Ausfallsicherheit und Disaster Recovery Fähigkeit
  - Kein Datenverlust bei Ausfall des Rechenzentrums → Wahrung der Datenkonsistenz
  - Steigerung der Performance im Bereich der Analytics, im speziellen der „freien“ Analysefähigkeit (unberechenbarer Workload)
  - Priorisierung von Diensten und Nutzergruppen (Abbildung der Serviceklassen operational BI vs. classic BI)
  - Einfache Skalierbarkeit
- **Die neue Architektur soll in der Lage sein, heutige und anstehende Probleme lösen zu können, ohne dass es notwendig ist, die Herausforderung an die Hardwareleistung anzupassen.**

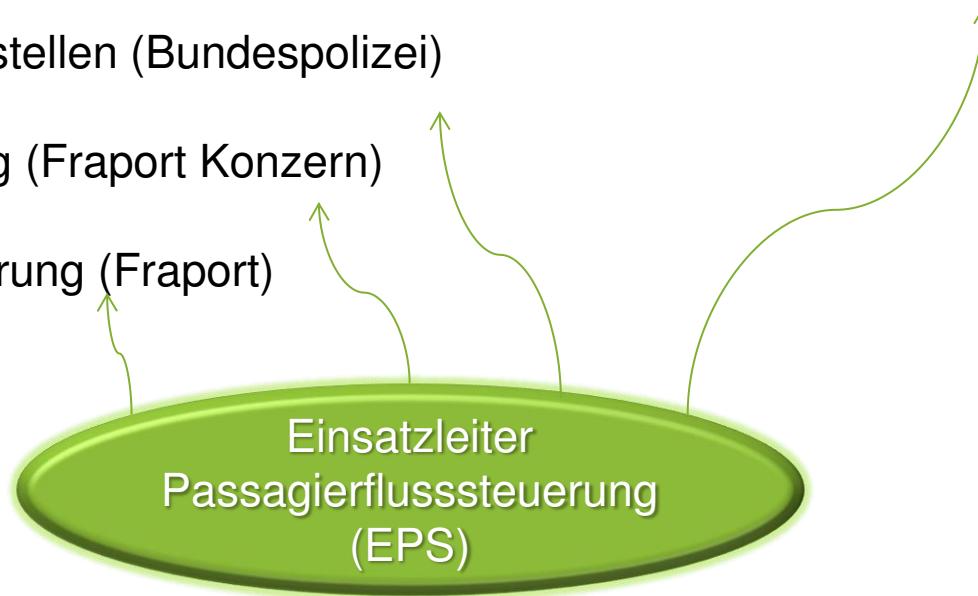
# **BACKUP**

## **Informationsfluss**

### Wer ist Nutzer der PFA?

Zur Zeit ca. 250 Nutzer in den Bereichen:

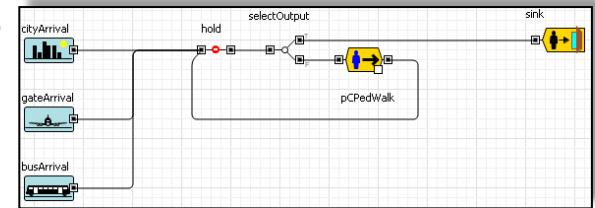
- § 5 Luftsicherheitskontrollen (Bundespolizei und Fraport Konzern)
- Grenzkontrollstellen (Bundespolizei)
- IVK Steuerung (Fraport Konzern)
- Betriebssteuerung (Fraport)



# **BACKUP**

## **Umsetzung (Simulation)**

- Aufbau einer Ereignis-Diskreten-Simulation unter Anwendung eines Pedestrian Models (agentenbasiertes Verhalten) und Verwendung des Terminallayouts als Navigationsgrundlage



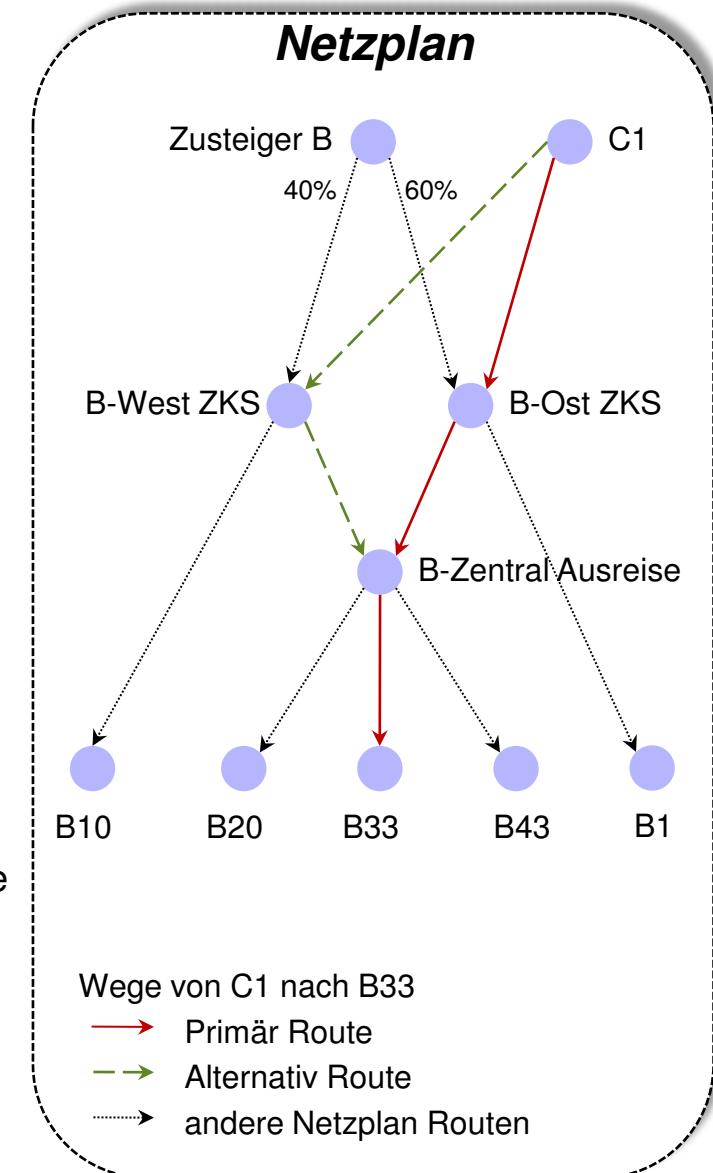
- Pedestrian Model in drei Detailstufen SocialForce Modell, Trajectory Modell, Simple Modell
- Ermittlung verschiedener Bedienzeitenverteilungen z.B. weibull, normal, uniform (manuelle Zählungen oder aus Systemen wie BCBP)
- Aufbau eines betrieblichen Netzplans (Terminal Layout) und einer Vektor-Liste zur Navigation der Passagiere innerhalb der Simulation
- Konzeption einer Simulations-Farm



## BACKUP

### Pax Navigation (Simulation)

- Als Grundlage dient das Terminal Layout aus dem sich implizit ein Netzplan ergibt
- Eine Vektorliste beschreibt die laut Beschilderung möglichen Routen
- Auf einer Kante (Fläche) bewegt sich der Passagier selbständig. Er interagiert mit seiner näheren Umgebung, geht anderen Passagieren aus dem Weg und sucht den kürzesten Weg zur nächsten Prozesstelle (Knoten). Er nutzt Elemente des Ebenwechsels nach eigenem ermessen.
- Nach jeder Prozesstelle entscheidet der Passagier erneut, welche Kante als nächstes zu benutzen ist. Stehen mehrere Kanten zur Auswahl, entscheidet eine Wahrscheinlichkeit über den Ausgang (z.B. ZUS-B → 40% / 60%)
- Fehlen Verbindungen (Kanten) entscheidet der Passagier autark und sucht sich den kürzesten Weg



# BIG Data im Gesundheitswesen – am Beispiel der Charité

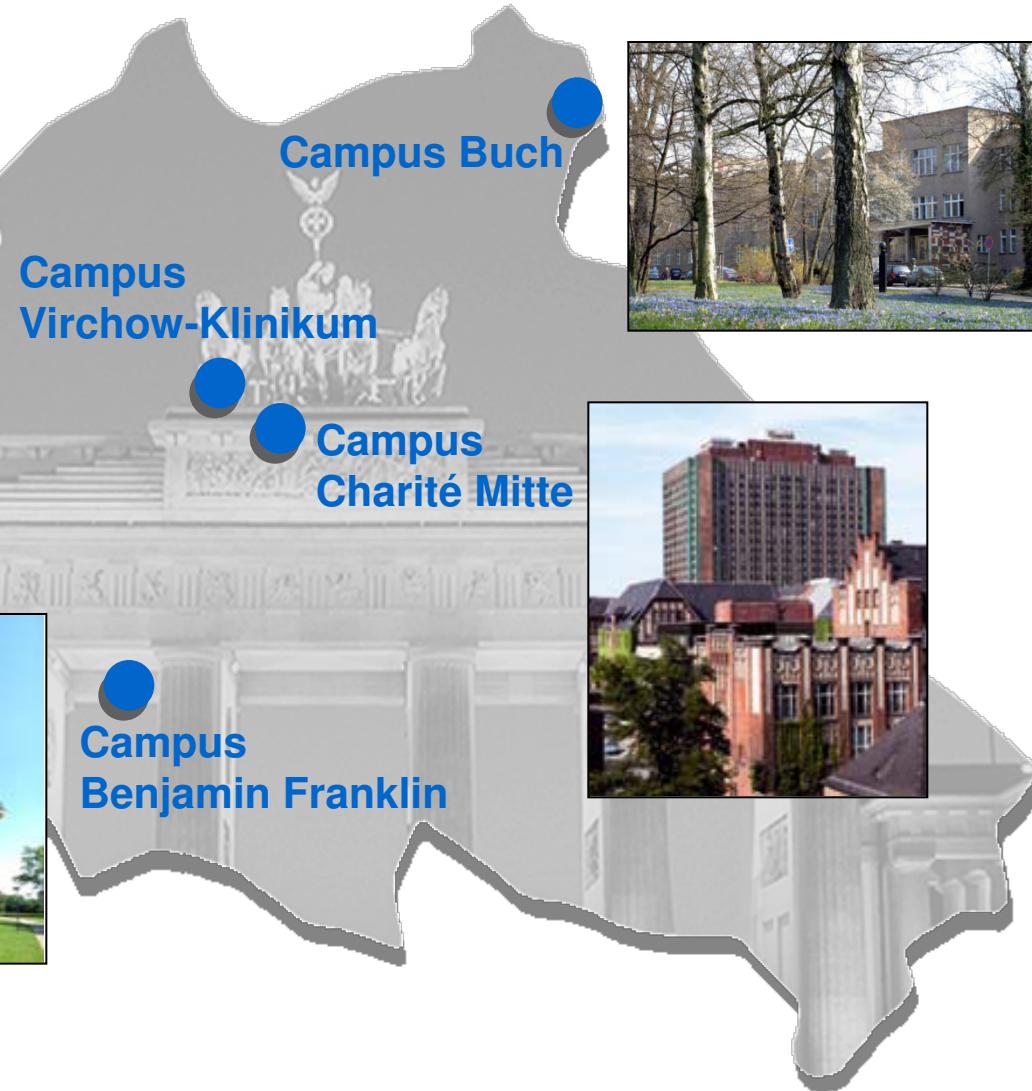


Martin Peuker

# Agenda

- Charité und Herausforderungen IT (BD)
- Integrierte Systemwelten
- Lösungsbeispiel:
  - Massendatenauswertungen
  - (Ansatz mobile Lösungen)

# Charité – Universitätsmedizin Berlin



# Kennzahlen

- ▶ **Vier** Berliner Standorte mit einer Nutzfläche von 607.200 m<sup>2</sup>
- ▶ **14.500** Mitarbeiter
  - ▶ darunter **3.736** Wissenschaftler und Ärzte,  
**4.065** Schwestern und Pfleger, **778** Verwaltungsmitarbeiter,  
**227** Professorinnen und Professoren **7,265** students
- ▶ **3,213** Betten mit einer durchschnittliche Verweildauer pro Fall **6.4 Tage**
- ▶ **1,3 Milliarde Euro** Jahresumsatz

Indicators	IST 2008	IST 2009	IST 2010	IST 2011	Δ 2011 - 2008	%
Casemix (Punkte)	181.890	186.786	192.910	198.229	16.339	9,0%
Stationäre Fälle	130.453	133.117	136.490	139.142	8.689	6,7%
Verweildauer (Tage)	7,40	6,67	6,56	6,41	-0,99	-13,4%
Ambulante Fälle	530.238	573.966	563.433	593.614	63.376	12,0%
Drittmittel (Mio. Euro)	126	131	149	158,2	32	26%

# Klinische Ziele

## (Massendaten Auszug)



- Unterstützung personalisierte Therapien
- Erforschung medizinisch-linguistischer Texterschließung (Arztbriefe, Befunde, Epikrisen, Verlegungsberichte...)
- Entwicklung kommerzieller Studiensoftware
- Sekundäre Datennutzung für Machbarkeitsanalysen klinischer Studien
- Sekundäre Datennutzung für Studienrekrutierung
- Vervollständigung von Case Report Forms (eCRFs) für klinische Studien
- Retrospektive Analysen  
(Qualitätssicherung SGB V 135a / 137c , Pharmakovigilanz, Data Mining, „eigene Forschung der Kliniken“)

*... Beitrag zum Senatsziel „Gesundheitsmetropole Berlin“ (BFG)*

# IT Herausforderungen (BD)

(Digitale Klinik)

- Viele Medienbrüche (noch immer papiergestützte Dokumentation notwendig?)
- Unstrukturierte Daten
- Anwenderorientierte Oberflächen.
- Daten müssen vollständig und in Echtzeit vorliegen.
  - Personalisierte Medizin
- Daten müssen überall (mobil) verfügbar sein.
- Datensicherheit muss durchgehend gewährleistet sein.

## **VI. Entscheidungsunterstützung** (wissensbasierte System-Funktionen)



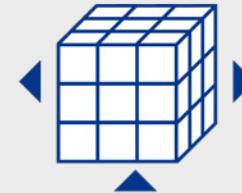
## **V. Automation (klinische Pfade)**



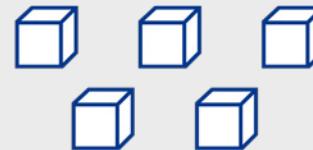
## **IV. Prozess- / Workfloworientierung**



## **III. Daten sammeln und bereitstellen**



## **II. Abteilungslösungen / Insellösungen**



## **I. Papiergestützte Datenerfassung**



# Realität!?

(Digitale Klinik)

Papiergestütztes Datenmanagement  
(für medizinische & administrative Prozesse)



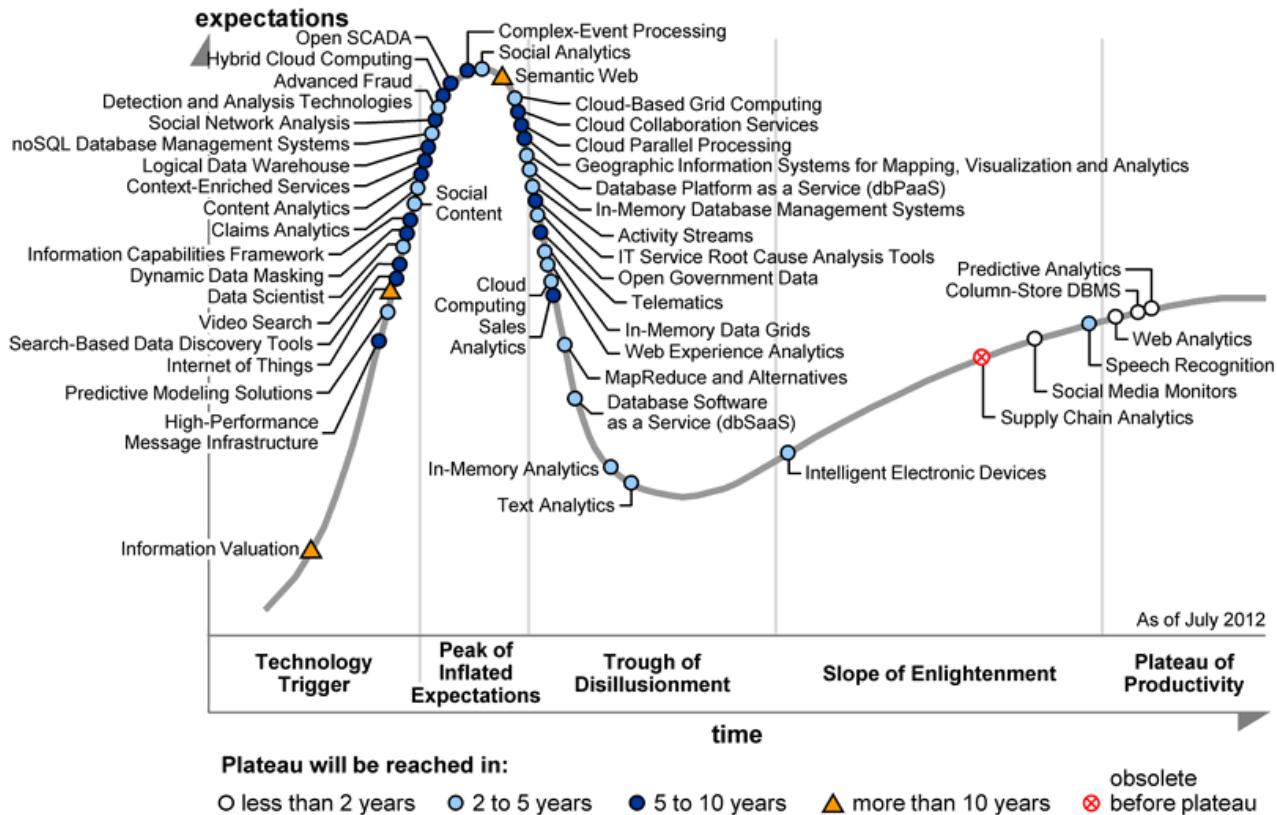
Workflow?

# BIG Data Technologieansatz (in Memory)



## SAP HANA InMemory Technologie

- Massendatenverarbeitung in Echtzeit (Vorhersage)
- Echtzeitsuche in Texten (auch unstrukturiert)
- Ziel Ausbau: Genetik und Systembiologie
  - IT Test am Bsp. „HANA Medical Explorer“ erfolgt
- Projekt zum SAP BW Upgrade auf SAP HANA (Standard)
- Enabler für diverse Anforderungen des klinischen Alltags und Forschung
  - „neue“ Qualität analytischer Anwendungen
- Basis für Funktionserweiterungen im Zusammenspiel mobiler und cloudbasierter Dienste

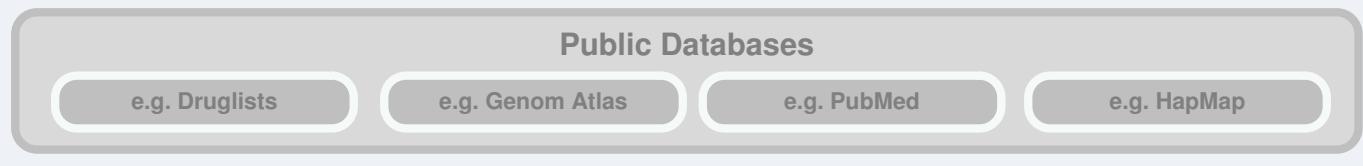
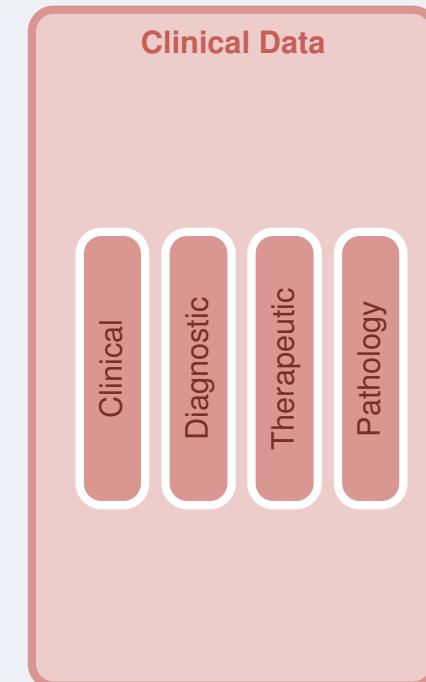
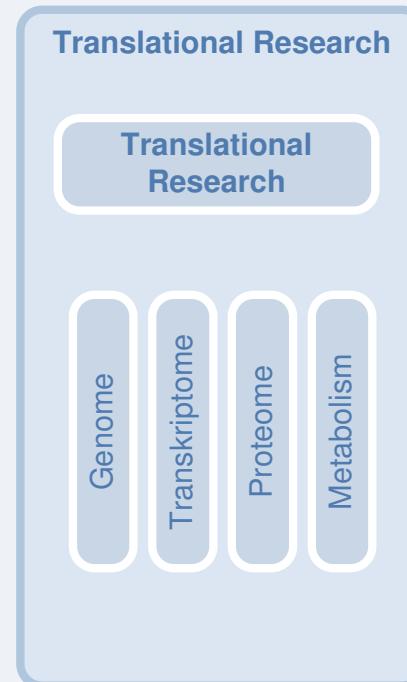
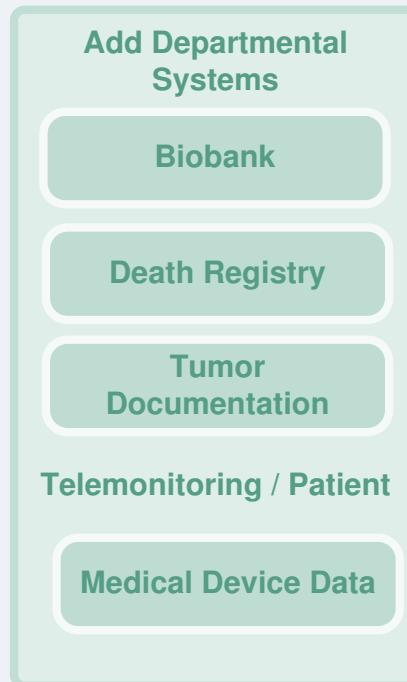


Source Gartner – BIG Data HypeCycle 2012

# IT architecture at Charité



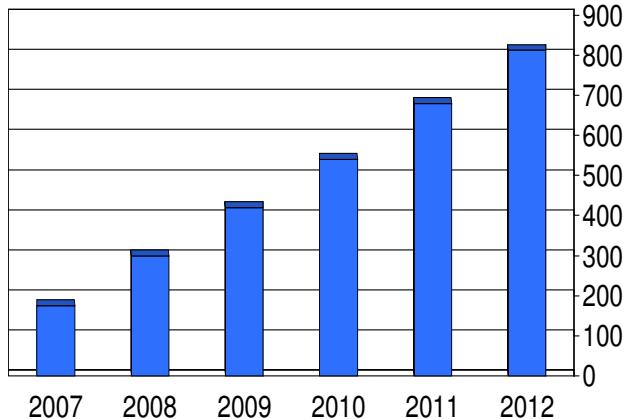
## User Access InMemory Applications



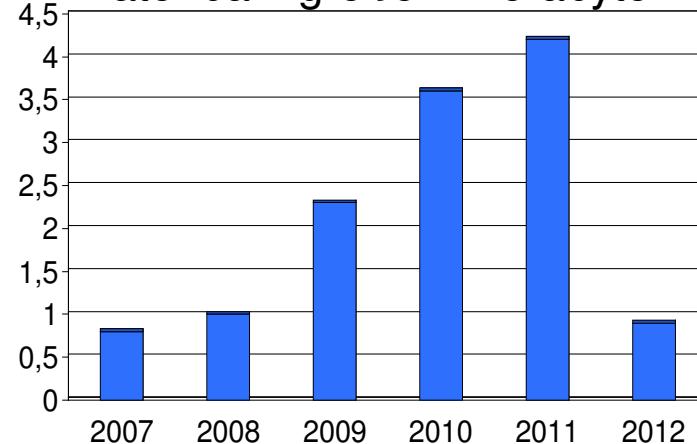
# Medizinisches DataWarehouse

Bsp. Charité

## Anzahl Nutzer DWH

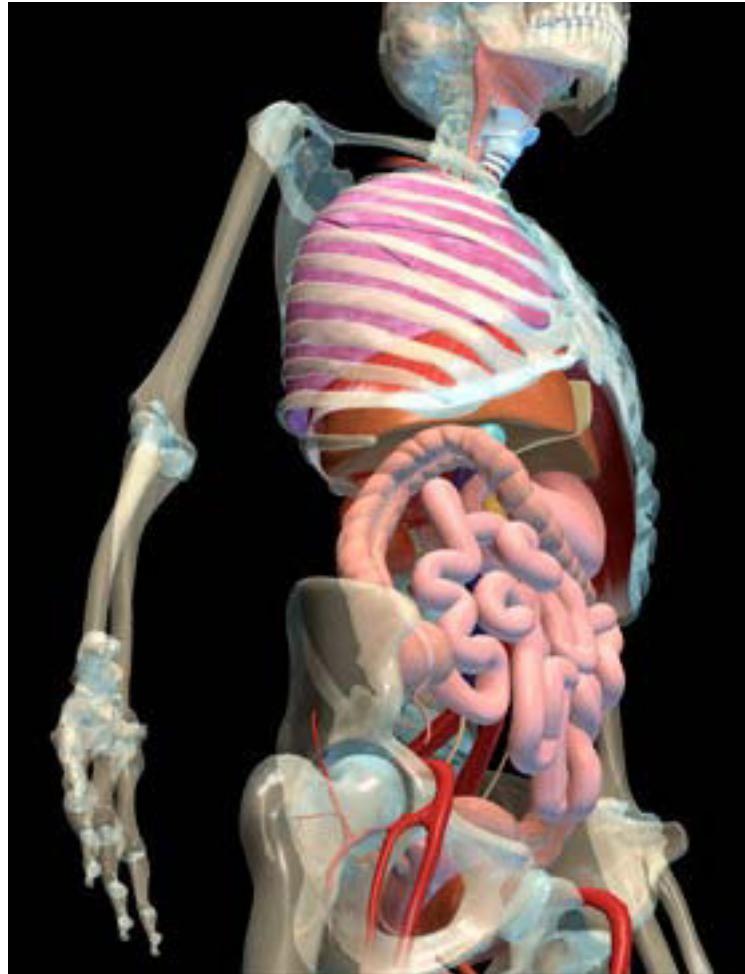


## Datenbankgröße in Terabyte



# Massendaten – Big Data

## Scenarios (for Research issues)



in addition to clinical data:

**Genome Data**

- 3.000.000 data points

**Transcriptome Data**

- 600.000 data points

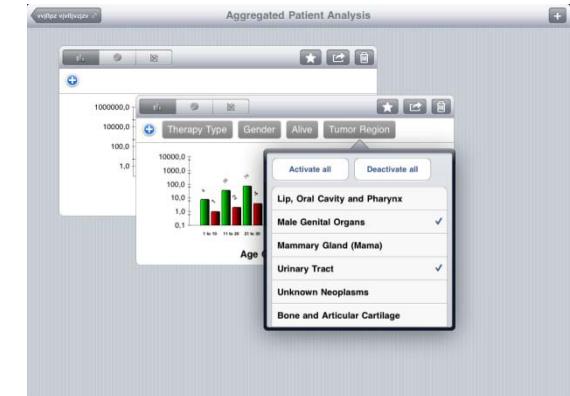
**Drug Screening**

- 1.000 data points

# Massendaten – Big Data

## Scenarios (for Research issues) - Medical Explorer

- ▶ In-Memory Technology as key-enabler for real-timeanalysis of tumor data in seconds instead of hours
- ▶ In-Memory enables join of third-party data to improve analysis results
- ▶ Information available at your fingertips: In-Memory Technology on mobile devices (iPad)



# Massendaten (InMemory) – Big Data

## Medical Explorer (for Research issues) -Fuzzy Search

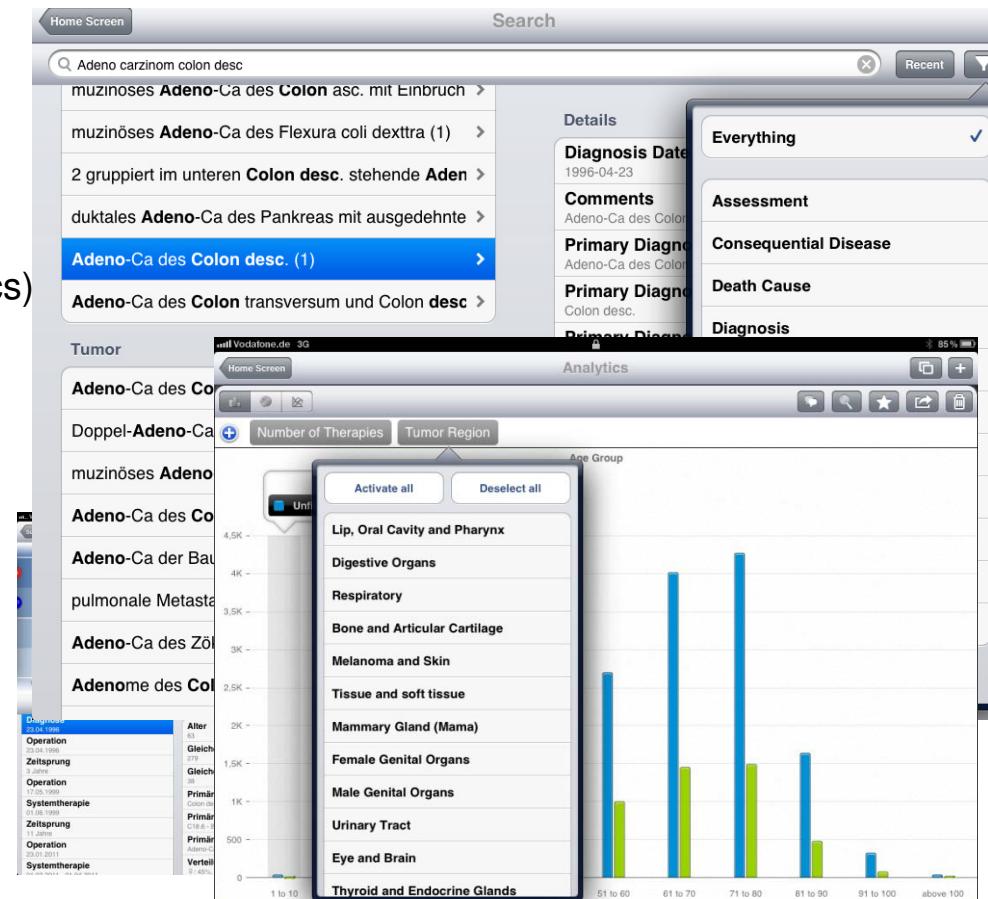


### Ad-hoc Analysis of heterogeneous tumor data for cancer research

- Medical records of decades of tens of thousands of patients
- Structured and unstructured data (records, time series, free text, etc.)

### Solution

- Integration into condensed but exhaustive view
- On-the-fly analyses (e.g. Kaplan-Meier estimation, cohort statistics)
- Attributes can be native, views, freetext-extracted, calculated

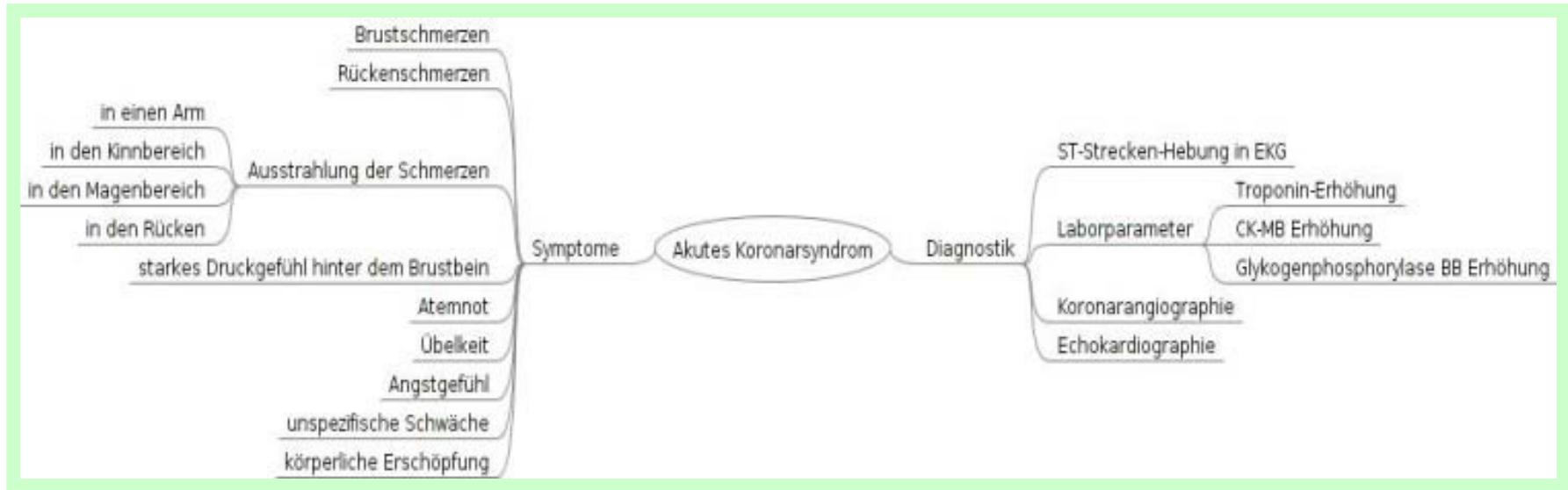


### Handling of unstructured data

- Any free text attribute is indexed
- Native support of typo-tolerant matching (using different string comparison methods)
- Native support of synonym tables (automatic retrieval of German/English records)

# BigData Herausforderungen Charité

## Erschließung von Wissen: Medizinisches Modellwissen (Ontologien) zur Aufdeckung von Zusammenhängen und Disambiguierung



Ontologieausschnitt Akutes Koronarsyndrom

# BigData Herausforderungen Charité

## Textbeispiel aus Aufnahmeprotokoll

### Zweiter Lösungsschritt: Morphosyntaktische Analyse und semantische Annotation

[Der Patient] [wird] [aus dem Heim] [vorgestellt] [wegen [einer [seit dem Vorabend] bestehenden Sehstörung], [apraktischen Fehlhandlungen] [sowie [intermittierender Wesensveränderung] [mit [zeitlicher Desorientierung] und [visuellen Verkennungen]]]].

543210: temporal  
543210: zeitlich

012345: Verwirrtheit  
012345: Desorientiertheit  
012345: Desorientierung



# BigData Herausforderungen Charité

- **Vielgestaltiger medizinischer Sprachgebrauch**  
(Insult, Apoplex, Schlaganfall, Stroke, Läsion Arteria cerebri media...)
- **Komplexer Skopus, Abkürzungen, Homonyme, Schreibfehler**  
(HWI, 5 mg 1-1-1, 5m, 5 m, Herzinfekt)
- **„Unscharfe Beschreibungen“**  
(flaue Flusssignale; mottenfraßähnliche, walnußgroße...)
- **Schwierige zeitliche Zuordnung der Symptome, Diagnosen und Maßnahmen**
- **Unterscheidung von akuten und chronischen Zuständen**
- **Gültigkeit und Wertigkeit von Befunden**
- **Datenschutz und Berechtigungen**

# BigData Ausblick Charité

- Anforderungen an bereits strukturierte Daten
  - Verbesserung der Datenverfügbarkeit bei Kassenverhandlungen, diverse BenchmarkSzenarios
- Verbesserte ERP Prozesse
  - Abrechnungs- und Mahnlaufperformance
  - Verbesserung der Qualitätskennzahlen (Realtime)
    - Bsp.: Echzeitverarbeitung Tracking OP Siebe
- “Echte Vorschungsunterstützung → Bsp. Tumorzentrum durch In-silico (machine-learning based)
- Datensupport für Pharmakooperationen

# Zusammenfassung / Quellen

## Literatur

- [Bit12] Bitkom: AK Big Data: Vorläufiges Programm 2012: in: [http://www.bitkom.org/files/documents/Programm\\_AK\\_Big\\_Data\\_2012.pdf](http://www.bitkom.org/files/documents/Programm_AK_Big_Data_2012.pdf); zugegriffen am: 14..4.2012.
- [Bra10] Bradley P, Kaplan J: Turning hospital data into dollars. Health Finance Management. 2010 Feb;64(2):64-8.
- [Bux11] Buxton, Ima: Echtzeitanalyse mit In-Memory: Mit iPad und ohne Handbuch unterwegs; in: Computerwoche vom 11.05.2011, [www.computerwoche.de](http://www.computerwoche.de), zugegriffen am 13.4.2012.
- [Cha10] Charité: SAP-Verfahrensdokumentation: Anlage zur Verfahrensmeldung SAP-03 „Grundsätze der Systemführung und der Zusammenarbeit“, Version 1.1, 14.6.201, Charité 2010.
- [Fer10] Ferranti JM, Langman MK, Tanaka D, McCall J, Ahmad A : Bridging the gap: leveraging business intelligence tools in support of patient safety and financial effectiveness. J Am Med Inform Assoc. 2010 Mar-Apr;17(2):136-43.
- [ISR11] IS-Report: SAP HANA im Einsatz bei der Charité Berlin, in: isreport, 9/2011, S. 28-29, [www.isreport.de](http://www.isreport.de), zugegriffen am 12.4.2012.
- [McC10] McCormack J : Number crunching. Predictive analytics helps Presbyterian Healthcare analyze financial data. Health Data Manag. 2010 Mar;18(3):99.
- [Nes12] Nessler, Susanne: Oncolyzer für Patientendaten – Mobile IT zur personalisierten Krebstherapie; in: Deutschlandradio Kultur, 27.02.2012, [www.dradio.de](http://www.dradio.de), zugegriffen am 12.4.2012.
- [TSy12] T-Systems: Whitepaper „SAP Innovationen“, 2012. T-Systems International GmbH, Frankfurt am Main, März 2012.
- [WHO10] World Health Organisation: World Health Report 2010. Health systems financing. The path to universal coverage.
- [ZPl10] Zeier A, Plattner H: In-Memory Data Management: An Inflection Point for Enterprise Applications, Springer 2010

# Fragen?

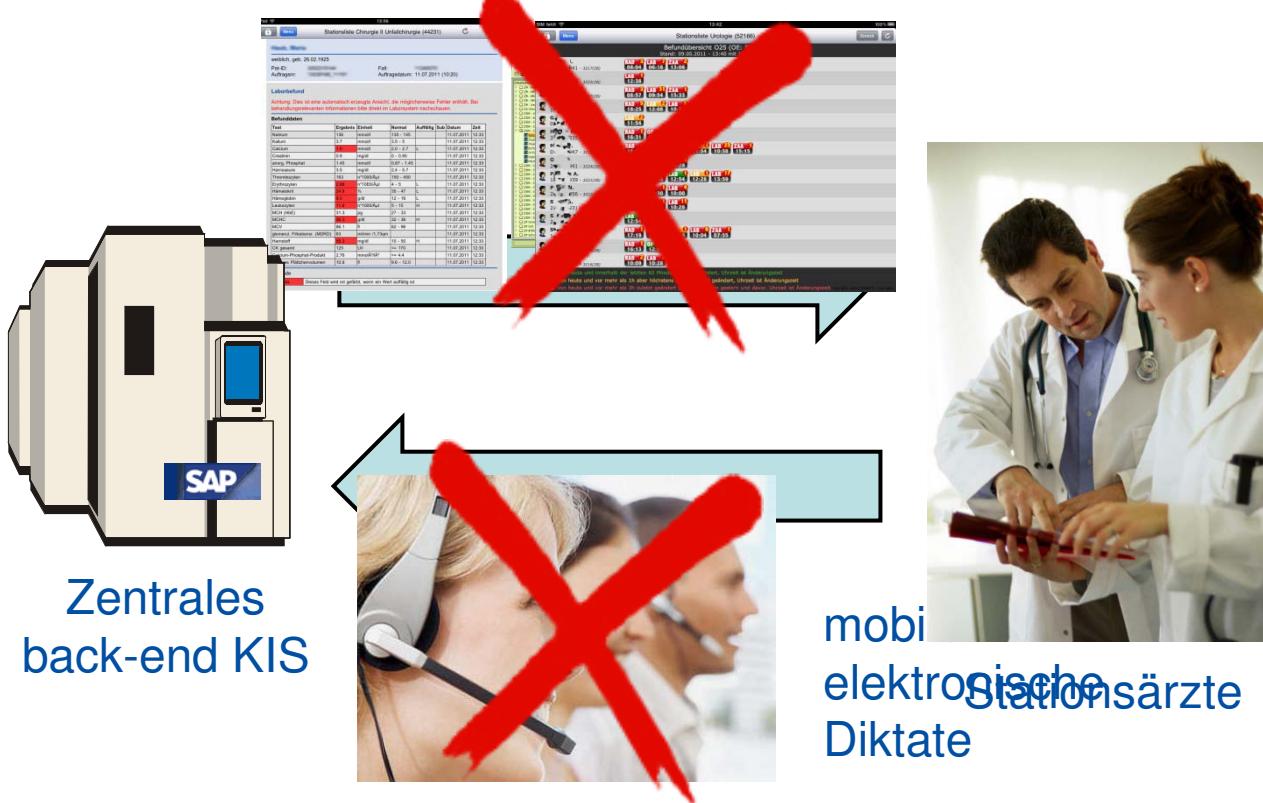
Martin Peuker

[martin.peuker@charite.de](mailto:martin.peuker@charite.de)

Charité Universitätsmedizin Berlin  
Charitéplatz 1  
10117 Berlin

# Neuer Ansatz – mittels sorgfältig ausgewählter IT-Komponenten

## Arbeitsplatzabhängige iPad-Applikationen



- Keine direkte Interaktion mit dem zentralen (back-end) KIS
- Mobile Komponenten zeigen (nur!) die relevanten Informationen
- Zentraler Pool medizinischer Assistenten, um die Ärzte von administrativen Tätigkeiten zu entlasten und Spracherkennung

# Bsp. ePA Mobil

**Meine Stationen - Station A**

**STATION A, RAUM 6**

- Michael Brown  
Fraktur des kleinen Fingers  
Station A, Raum 6
- Geronimo Oley  
Trigeminusneuralgie  
Station A, Raum 6

**STATION A, RAUM 8**

- Steve Miller  
Trigeminusneuralgie  
Station A, Raum 8
- James Donalds  
Fraktur des kleinen Fingers  
Station A, Raum 8
- John Bush  
Trigeminusneuralgie  
Station A, Raum 8
- Steve Erin  
Fraktur des kleinen Fingers  
Station A, Raum 8

**Geronimo Oley ( ♂ | 45 | 01.01.1967 )**  
Trigeminusneuralgie  
Station A, Raum 6 / Aufnahmedatum 10.8.2011 (10. Tag)

Allergien/Risiken  
Für den ausgewählten Zeitraum existieren keine Daten  
Allergien

Medikamentenunverträglichkeit, Sturzgefahr  
Risiken

Dokumente

Begleitbrief Einweisung 20.08.11 10:30 Aktiv

Bilder

Röntgen Schädel 11.08.11 12:01 Aktiv

CT Schädel 11.08.11 13:08 Aktiv

Klinische Aufträge

CT Schädel 11.08.11 13:08 Erfolgt

Röntgen Schädel 11.08.11 12:01 Erfolgt

**John Bush**

John Bush ( ♂ | 34 | 01.04.1978 )  
Trigeminusneuralgie  
Station A, Raum 8 / Aufnahmedatum 10.8.2011 (10. Tag)

Kurven

Dokumente

Bilder

Labor

Klinische Aufträge

**Kodierung**

Patientendaten

Verlaufsnotizen

Klinische Aufgaben

SAP EMR Info

Probleme

FPI Gesichtsschmerz 11.08.11

ED1 Erschöpfungszustand 11.08.11

Diagnosen

G50.0 (Haupt) Trigeminusneuralgie 11.08.11

Z73.0 Probleme mit Bezug auf Schwierigkeiten bei der Lebensbewältigung 11.08.11

I10 Essentielle Hypertonie, nicht näher bezeichnet 10.08.11

J32.8 Sonstige chronische Sinusitis 10.08.11

Prozeduren

1-699.0 Endoskopie der Nasennebenhöhlen 11.08.11

# Bsp. ePA Mobil

SIM fehlt 09:44 99%

**John Bush**

Trigeminusneuralgie

Station A, Raum 8 / Aufnahmedatum 10.8.2011 (10. Tag)

**Kurven**

**Dokumente**

**Bilder**

**Labor**

**Klinische Aufträge**

**Kodierung**

**Patientendaten**

**Verlaufsnotizen**

**Klinische Aufgaben**

**SAP EMR Info**

**BLUTGASANALYSE**

	11.08.11 23:01	12.08.11 13:01	13.08.11 12:01	14.08.11 12:30
Basenabweichung (-2,5/+2,5 mmol/l)	+1,2	+2,8	-1,2	+1,0
Sauerstoffsättigung (93/99 %)	94,6	97,2	95,0	92,7
pCO <sub>2</sub> (4,7/5,7 kPa)	5,8	5,3	6,0	6,5
ph (7,36/7,44)	7,42	7,38	8,1	7,37
Sauerstoff (9/13,9 kPa)	9,0	8,0	7,38	9,2
Standard HCO <sub>3</sub> (20/28 mmol/l)	25,2	25,5	27,6	27,0

**ELEKTROLYTE**

Natrium (133-148 mmol/l)	135,2	142,0	145,5
Kalium (3,5-5,0 mmol/L)	5,0	3,0	4,1

**HAEMATOLOGIE**

Erythrozyten (4,15-4,90 Mio/L)	4,2	4,62	4,51
--------------------------------	-----	------	------

**HÄMATOLOGIE**

Leukozyten (4,3-10,8 billion/L)	6	6,5	7,4
---------------------------------	---	-----	-----



SIM fehlt 09:44 99%

**John Bush**

Trigeminusneuralgie

Station A, Raum 8 / Aufnahmedatum 10.8.2011 (10. Tag)

**Kurven**

**Dokumente**

**Bilder**

**Labor**

**Klinische Aufträge**

**Kodierung**

**Patientendaten**

**Verlaufsnotizen**

**Klinische Aufgaben**

**SAP EMR Info**

**Filtern nach**

**Notiz - Behandelnder Arzt**  
Michael Schneider, Arzt  
17.08.11 10:30  
Ernährung von John Bush sollte umgestellt werden - bitte die letzte Anordnung beachten!

**Notiz - Behandelnder Arzt**  
Michael Schneider, Arzt  
16.08.11 10:30  
John Bush fühlt sich heute besser, lediglich noch etwas geschwächt.

**Notiz - Pflege**  
Melanie Heart, Krankenschwester  
13.08.11 19:20  
John Bush fühlt sich heute sehr schwach und klagt über Kopfschmerzen und wirkt unruhig. Ich habe Dr. Miller gebeten, am Abend nach ihm zu sehen.

**Notiz - Pflege**  
Melanie Heart, Krankenschwester  
11.08.11 12:01  
Röntgen Schädel  
Röntgen Schädel (1 Bild)  
Röntgen Neurologie  
Dr. med. Stefan Hofstetter, (06221) 234 3245 (CR)

SIM fehlt 09:44 99%

**Meine Stationen - Station A**

**John Bush**

Trigeminusneuralgie

Station A, Raum 8 / Aufnahmedatum 10.8.2011 (10. Tag)

**Kurven**

**Dokumente**

**Bilder**

**Labor**

**Klinische Aufträge**

**Kodierung**

**Patientendaten**

**Verlaufsnotizen**

**Klinische Aufgaben**

**SAP EMR Info**

**Bearbeiten**

**Erstellte Alben**

**Album anlegen**

**1 Bild** Aktiv

**11.08.11 12:01**

Röntgen Schädel  
Röntgen Schädel (1 Bild)  
Röntgen Neurologie  
Dr. med. Stefan Hofstetter, (06221) 234 3245 (CR)

# What is big data?

---

- Quantity of data available
- Relevance
- Timely
- Complete & Accurate
- Is available when a decision needs to be made

# Big Data as new paradigm in the Financial Industry / Scoring

 **Kreditech**  
*BIG DATA Scoring*



*featured on*



TechCrunch

VentureBeat

Forbes

CNN

YAHOO!  
FINANCE

SPIEGEL  
ONLINE

Frankfurter  
Allgemeine  
ZEITUNG FÜR FRANKFURT

1. Big Data In Financial Industry
2. Big Data As Scoring Paradigm
3. Kreditech

# What is big data?

900% additional data per year in comparison to all data in human history.



Total data volume from first data ever recorded till  
**2010: 0,18 Zettabyte**  
**or 180.000.000.000 Gigabyte**

Data volume from  
2010 to 2011:  
**1,62 Zettabyte**

Cheaper Computational Power:

- Cloud computing
- In-memory databases

Smarter Algorithms:

- Data aggregation (map reduce)
- Machine Learning (artificial intelligence)

# How is big data being used?

Billion of records on private individuals



% of data used



2

FBI

Credit Bureaus



3%

Financial Industry

**EUROFORUM**  
Quality in Business Information

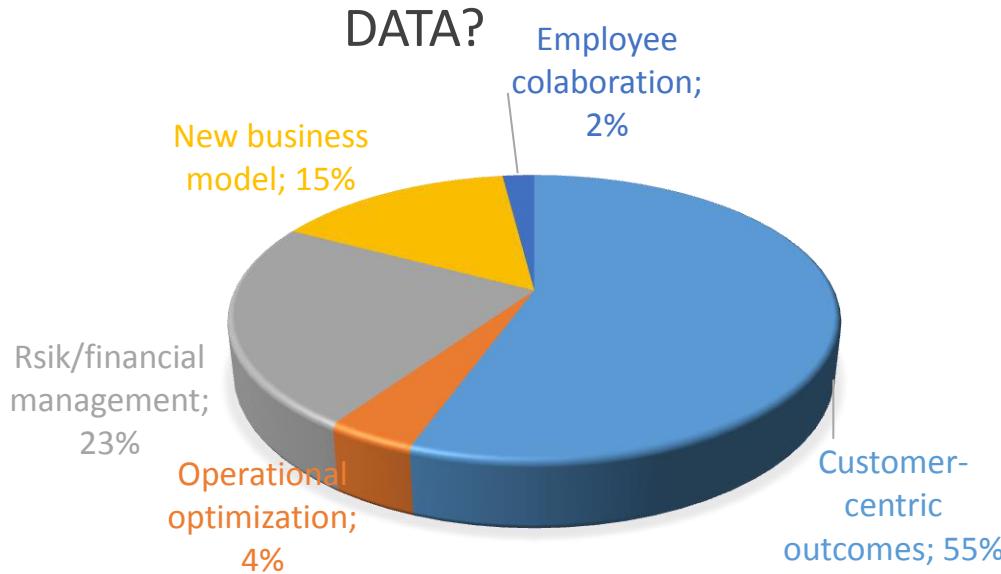
5

Source: Analytics: The real-world use of big data, a collaborative research study by the IBM Institute for Business Value and the Said Business School at the University of Oxford.

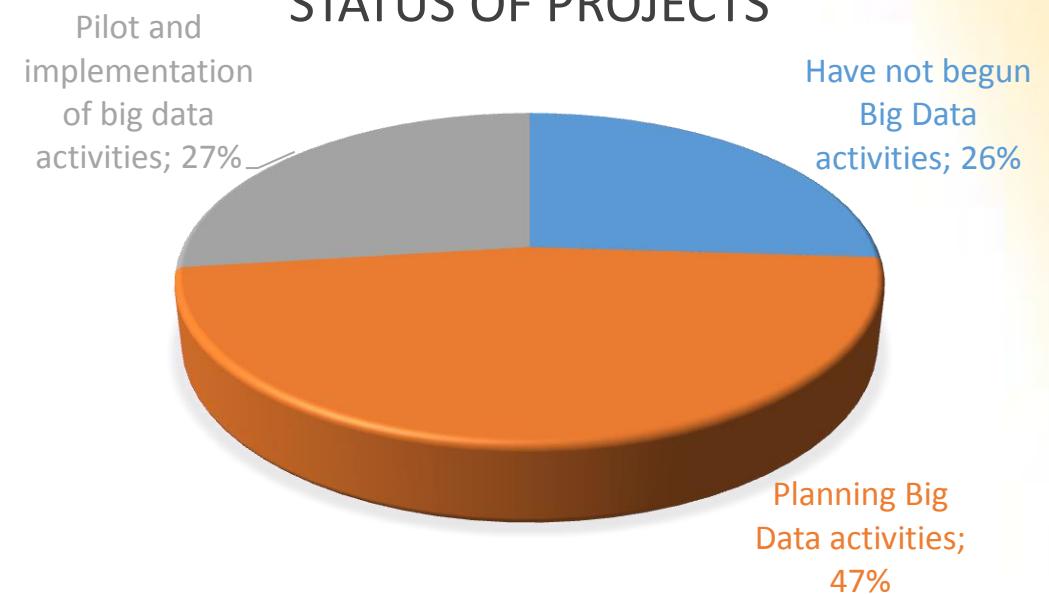
27.11.2013

# How do banks use big data?

## HOW DO BANKS SEE POTENTIAL IN BIG DATA?



## STATUS OF PROJECTS



*Big Data offers much more, than raw information. It's the foundation of new strategies.*

# Current Chances & Challenges

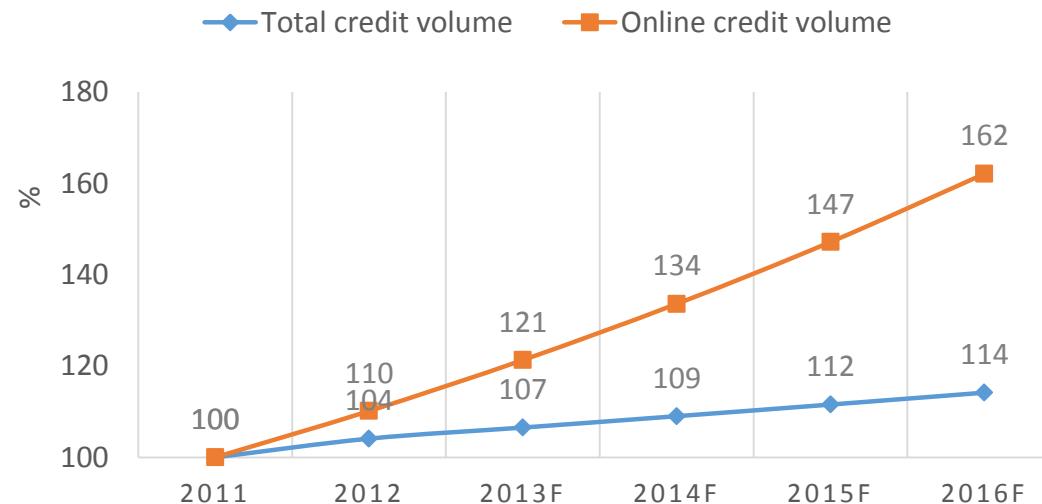
Chances	Challenges
<ul style="list-style-type: none"><li>- Generate more data (by Big Data Technologies)</li><li>- Analyze more data (e.g. Zalando vs. Shore store)</li><li>- Faster / Leaner/ Cheaper processes</li><li>- New revenue streams</li><li>- Developing innovative products</li><li>- Custom tailored solutions and offers</li></ul>	<ul style="list-style-type: none"><li>- Structural Change</li><li>- Channel-switch from Offline to Online</li><li>- Security &amp; Data protection</li><li>- Regulation</li><li>- Cannibalization of „old“ business models</li><li>- Technological Resources</li></ul>



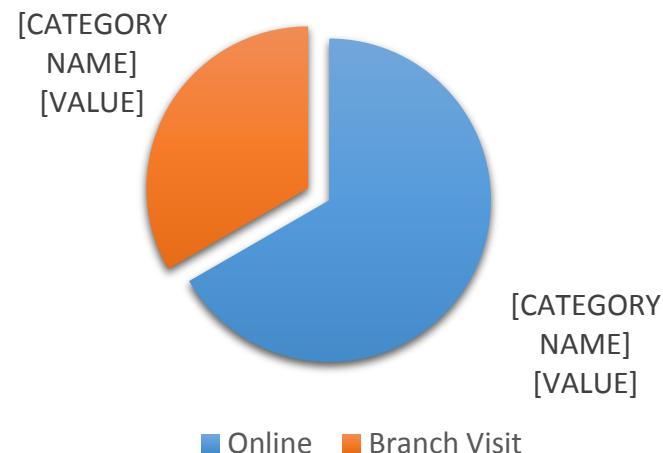
Source: Bigstockphoto

# Window of Opportunity: High-Street Banks will Lose Dominance of Lending Market

Exemplary: Consumer credit market development in the US (the world's biggest consumer credit market)



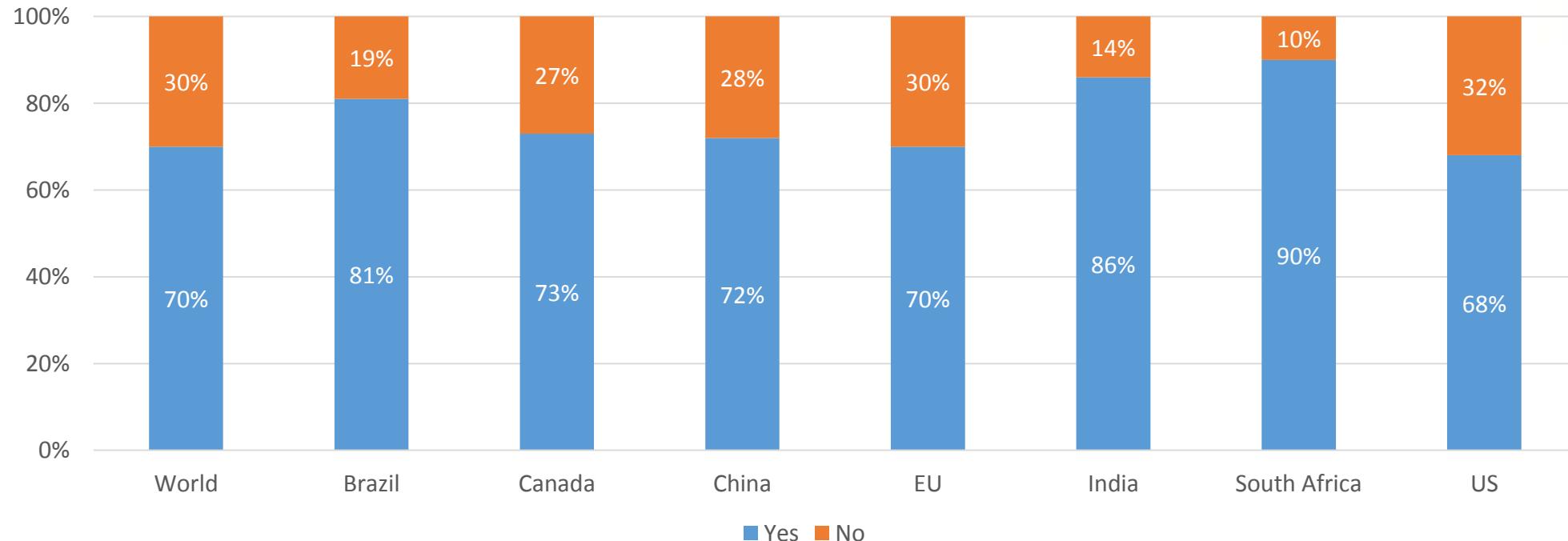
Customers' preferred channels for banking activities worldwide



*Online loans are growing 500% faster than traditional loans.  
High potential not only in emerging markets but also developed markets.*

# What does this mean for the industry

Would You be Willing to Provide Your Bank with More Information About  
Yourself if it Helped to Deliver a Better Service?



Source: "The customer takes control" - Global Consumer Banking Survey 2012 by E&Y;  
Consumer Credit - G.19 report by Federal Reserve (US)

27.11.2013

**EUROFORUM**  
Quality in Business Information

# What does this mean for the industry

3 possible end games



Banks get disrupted / shift in supply chain and market shares

Banks as innovators and innovation driven entities

Banks as gatekeepers / Infrastructure provides

*Analogy: Print Industry*



*Analogy: Photography Industry*



*Banks become the „itunes“ for financial services*



1. Big Data In Financial Industry
2. Big Data As Scoring Paradigm
3. Kreditech

# How can we use Big Data for Scoring?

## Using Big Data

- ✓ Relevant to the decision
- ✓ Complete & Accurate
- ✓ Timely
- ✓ Available when a decision is being made

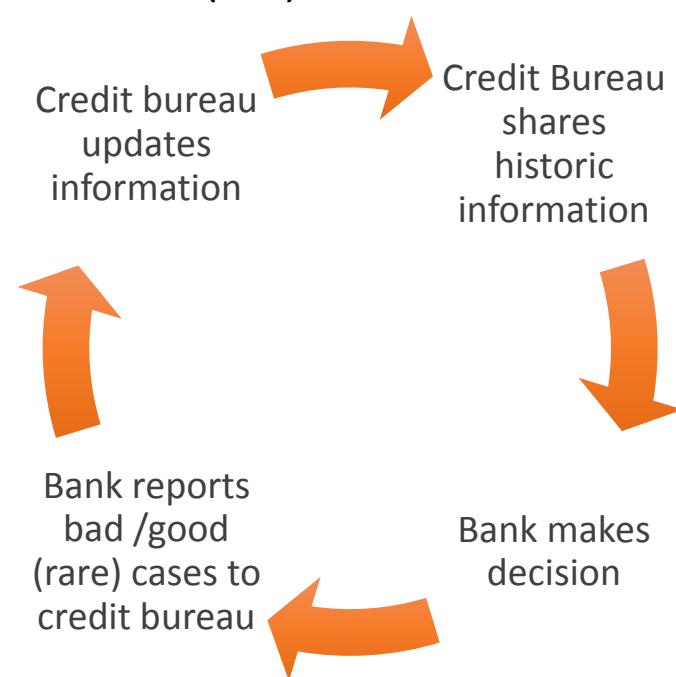
## Credit Bureau use of Big Data

- ✗ No access to really useful info.
- ✗ A lot of missing information
- ✗ Old and out of date
- ✗ Information not collected when decision is made

# 5bn customers don't have a credit score



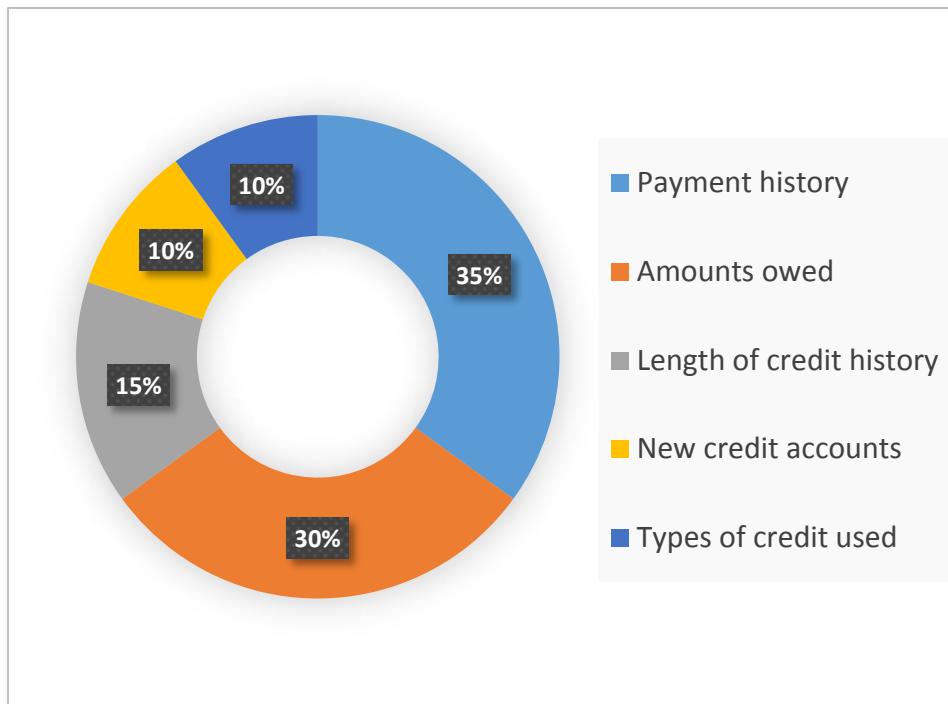
## How Banks (still) make credit decisions



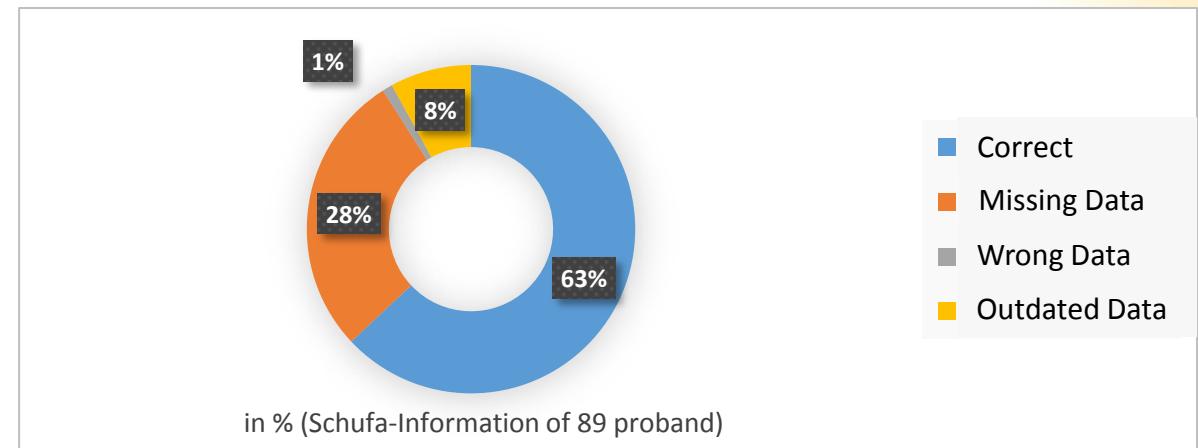
*Biggest market potential lies in emerging markets with no / very poor credit bureau data availability.*

# GIGO Model – Garbage in, garbage out

What makes up a FICO credit score?

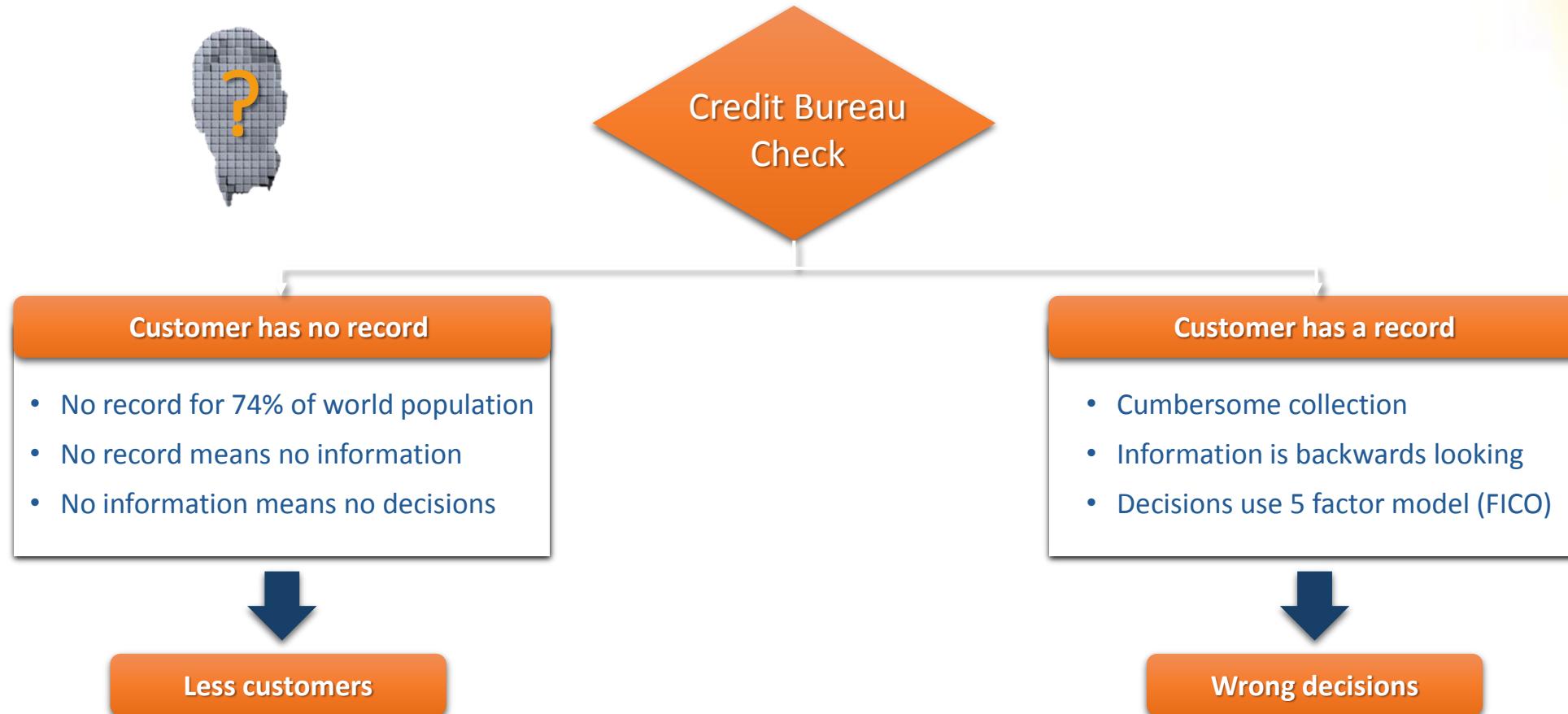


Germany: Insufficient Schufa Information

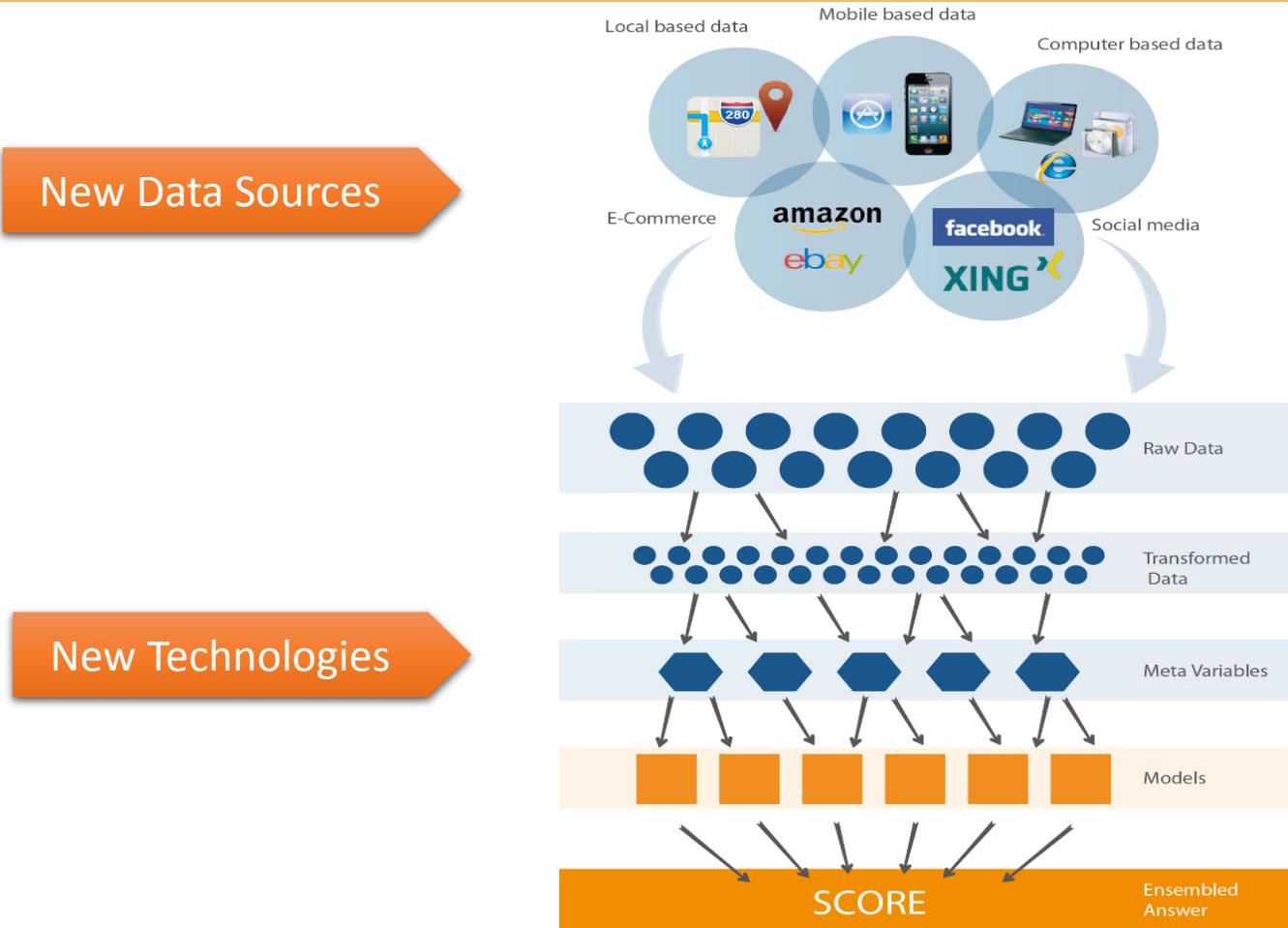


- Backward looking
- Clustered into „risk groups“
- No previous credit means you're less creditworthy

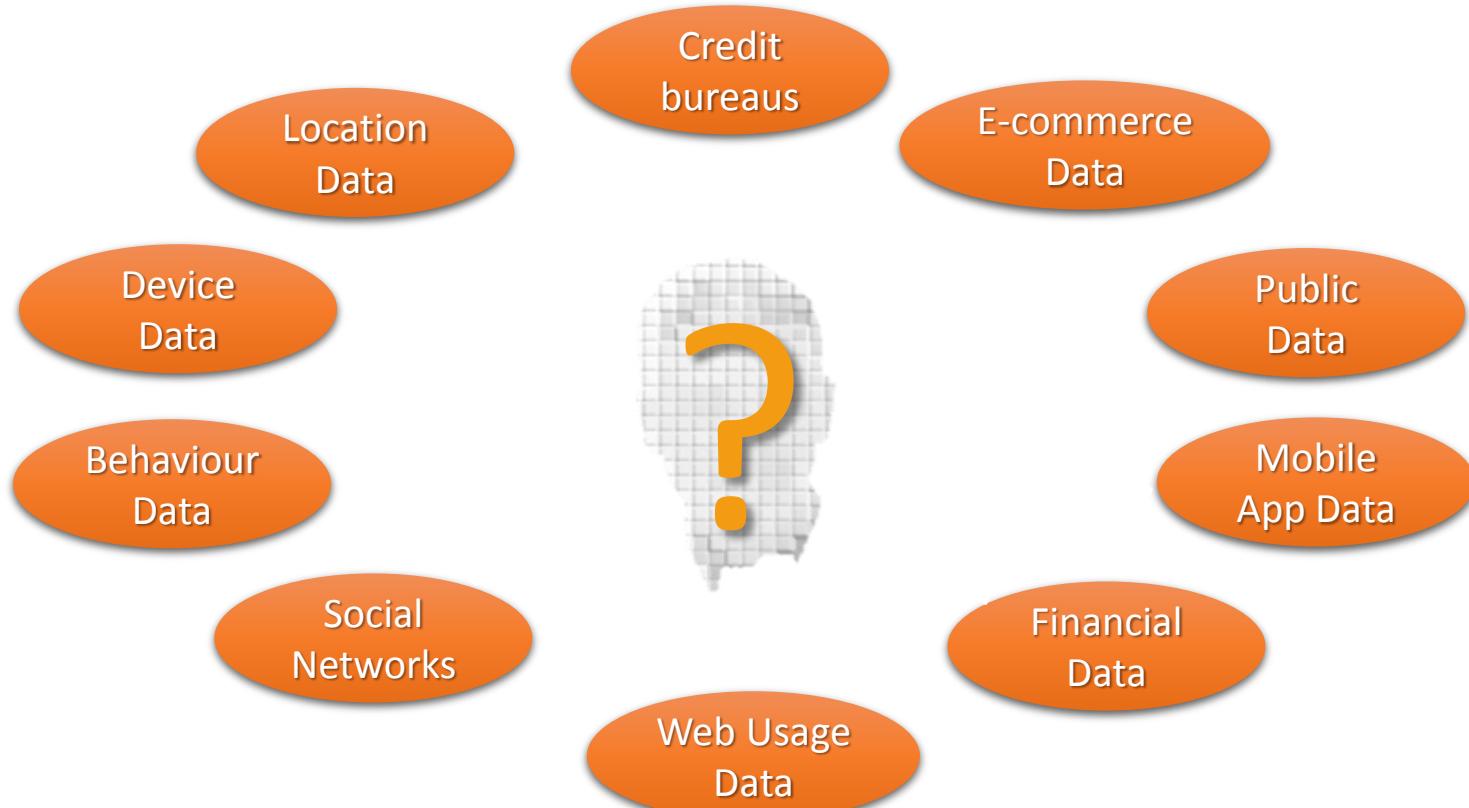
# Since Current Solutions are Insufficient



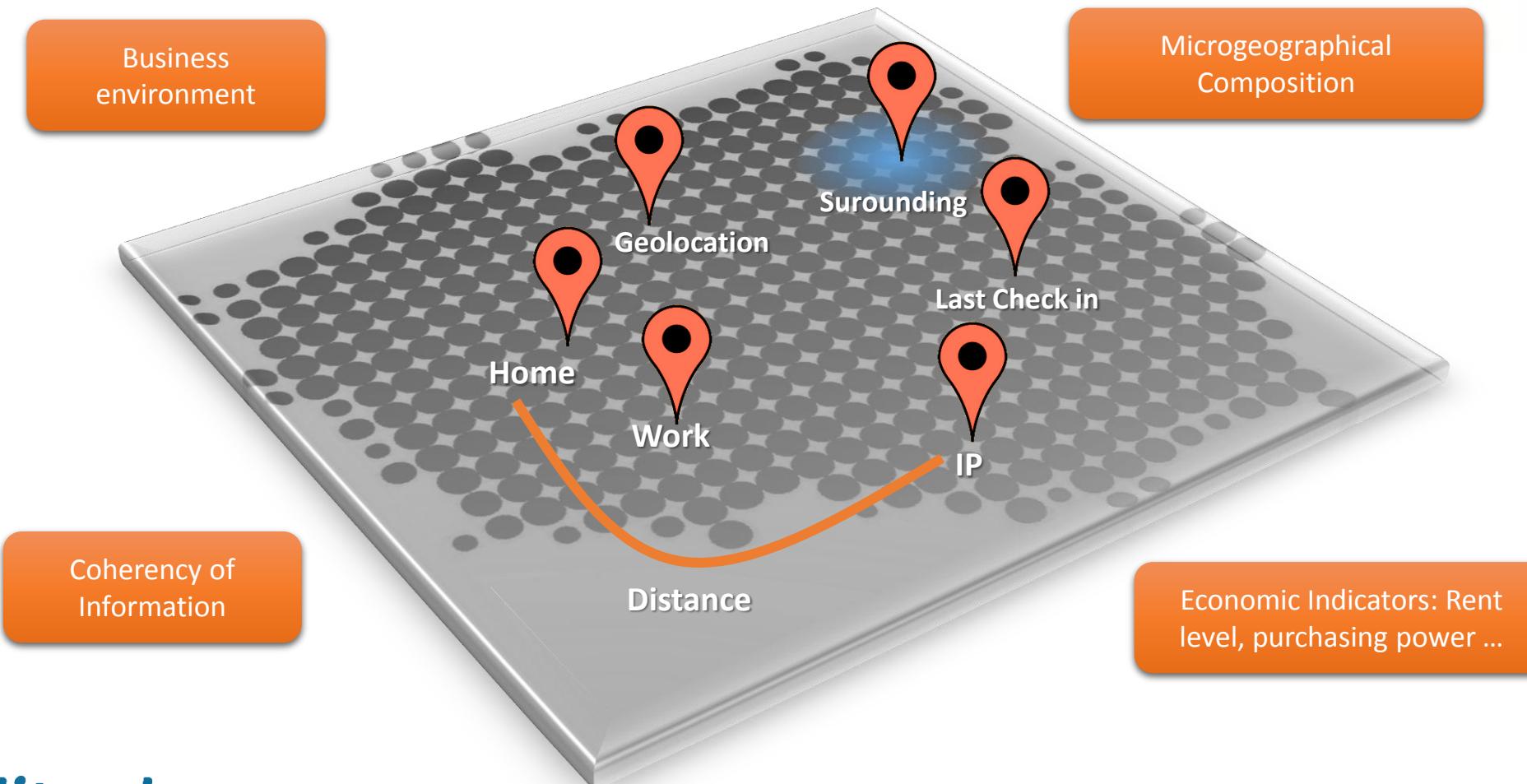
# Why not use online data – country and credit bureau independent?



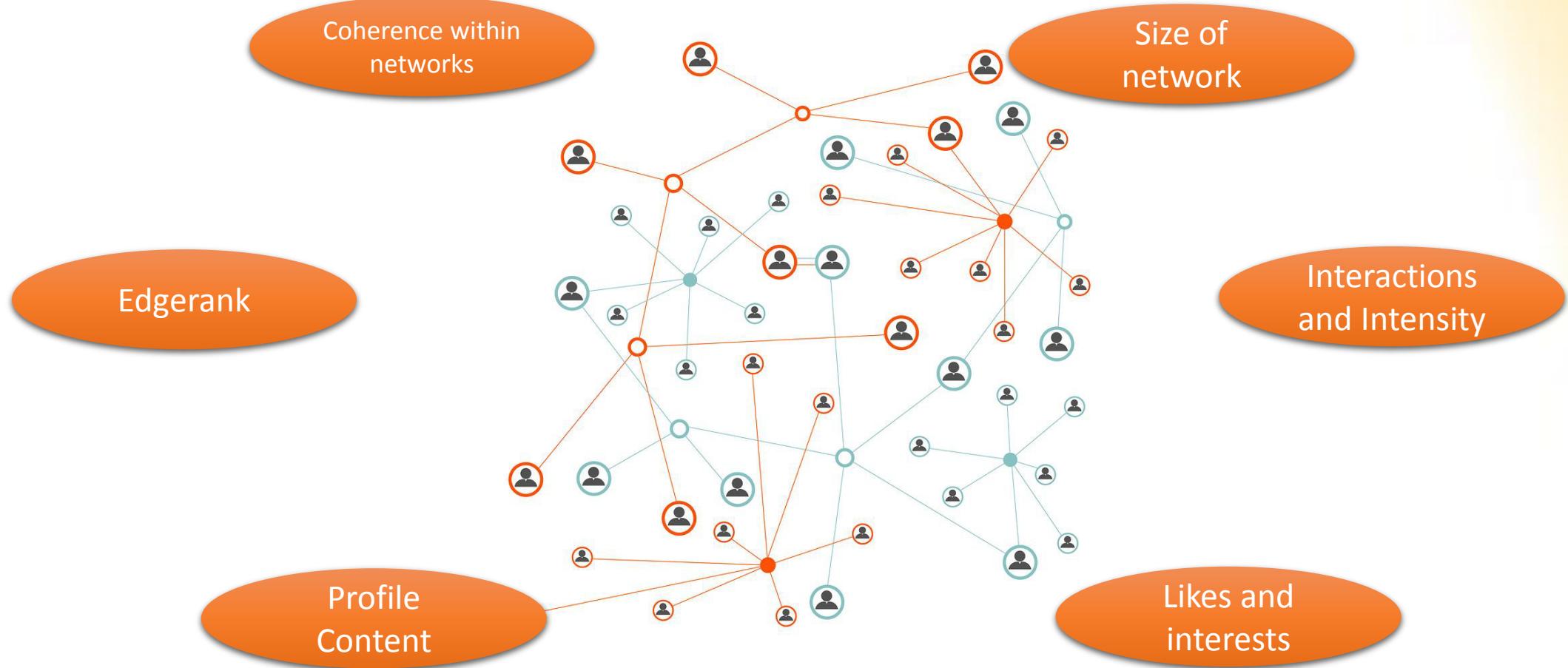
# Get more data!



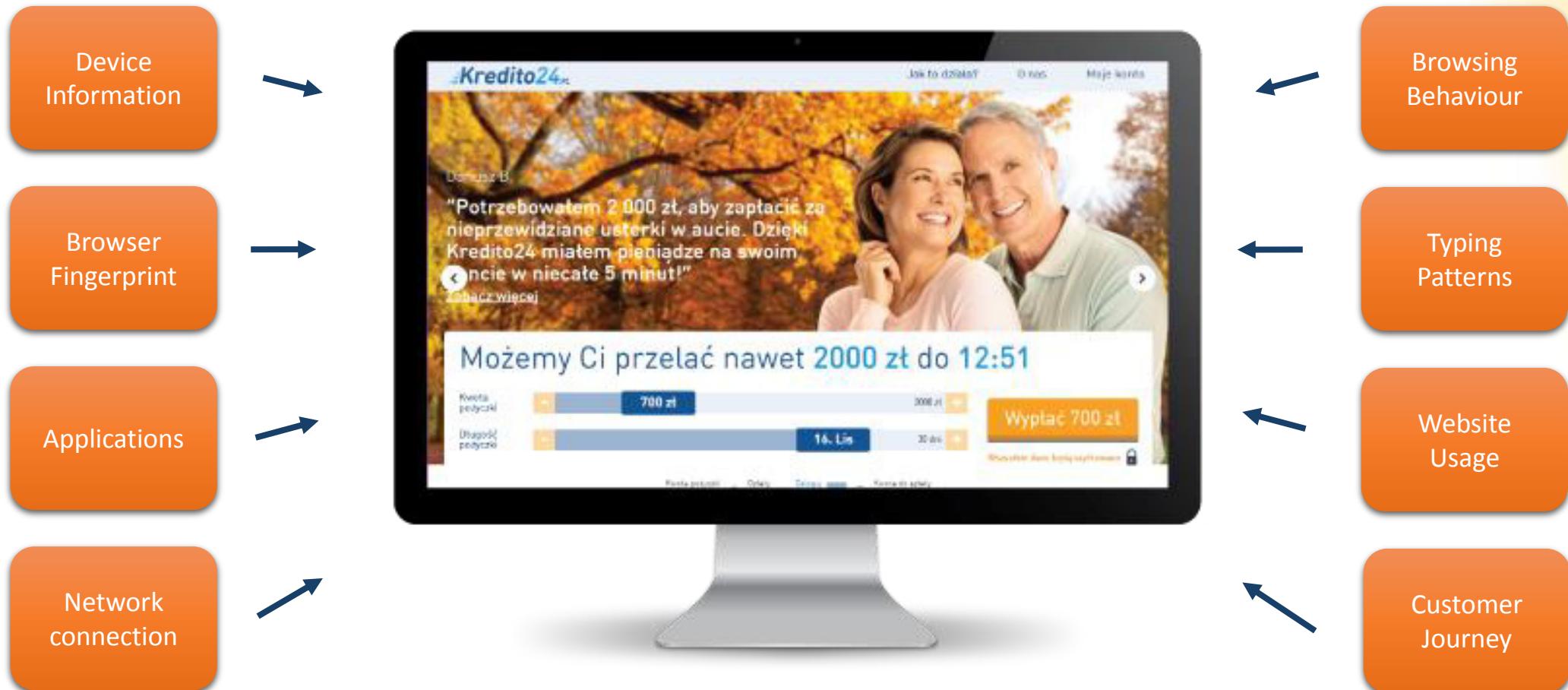
# Location and Environmental Data



# Social Network Data



# Device and Behavioural Data



# More Data is better.



*The big data approach can be seen as mosaic picture: the individual datapoint only has marginal influence*

1. Big Data In Financial Industry
2. Big Data As Scoring Paradigm
3. Kreditech

# Automated Microloans Based on Scoring Tech

## 1. Fast, convenient microloans

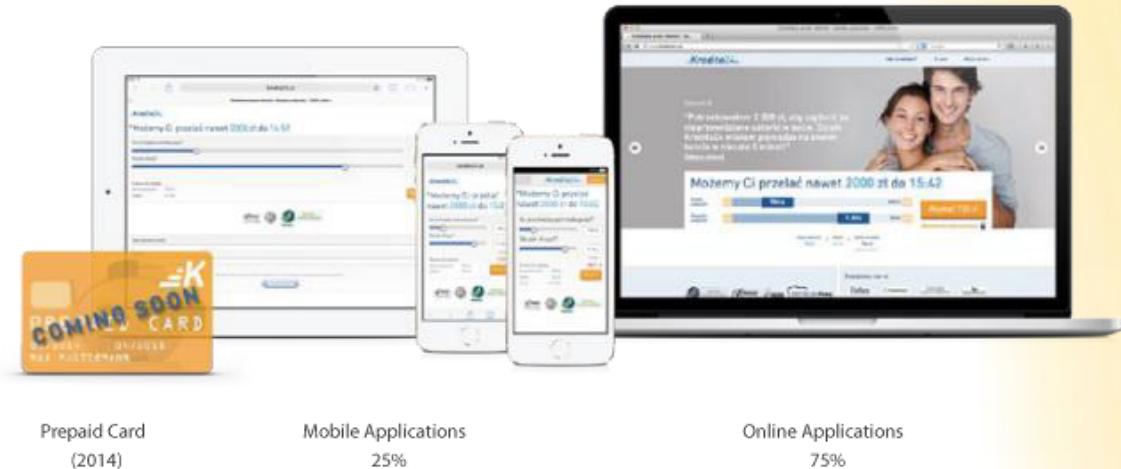
- 7 – 45 Days; installments recently launched
- 50 – 2.500 EUR (in local currency)
- Maximum of transparency, convenience and speed

## 2. High customer value

- 24/7 open (no credit bureaus needed)
- Instant decision, real time pay-out (in 10 min on bank account)
- No signature, scans / docs, calls etc. required

## 3. Automated, scalable operations

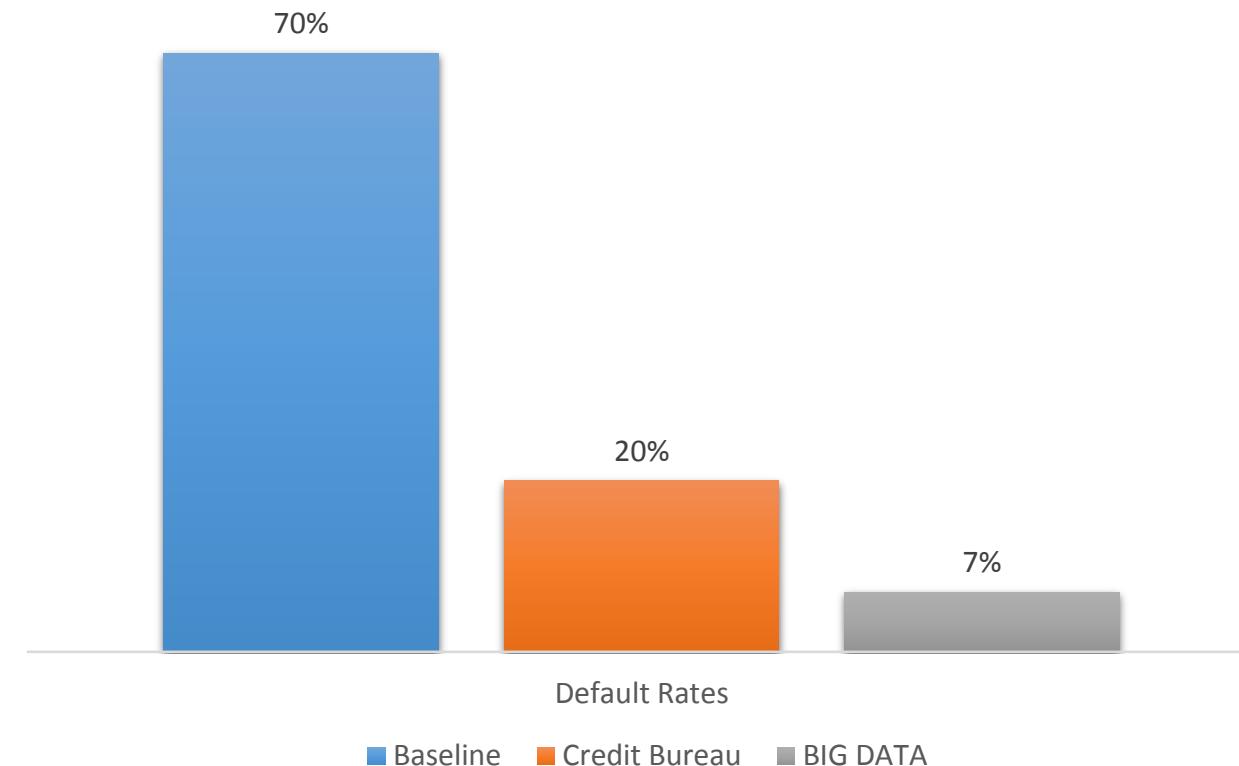
- No variable cost besides external query costs and transfer costs
- Scalability (revenue grew by 150 % to 7-digit p.a. with same headcount / fixed costs)
- Minimized errors and 24/7 operations



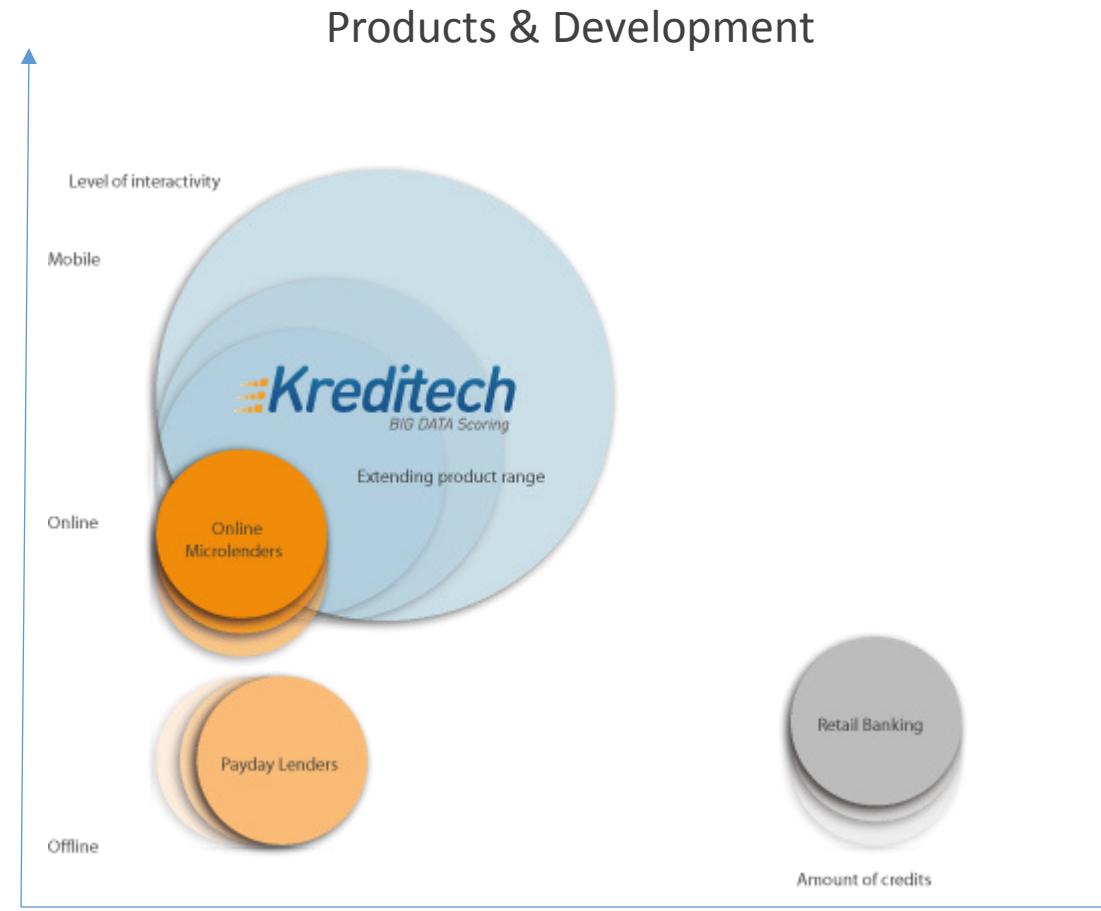
*Online loans enable maximum learning (volume) at minimum capital risk (low amount) in minimum time (short duration).*

# BIG DATA vs. Credit Bureau Data.

Default rates Poland 2013



# BIG DATA vs. Credit Bureau Data.



# Profitable, Automated, Scalable Operations

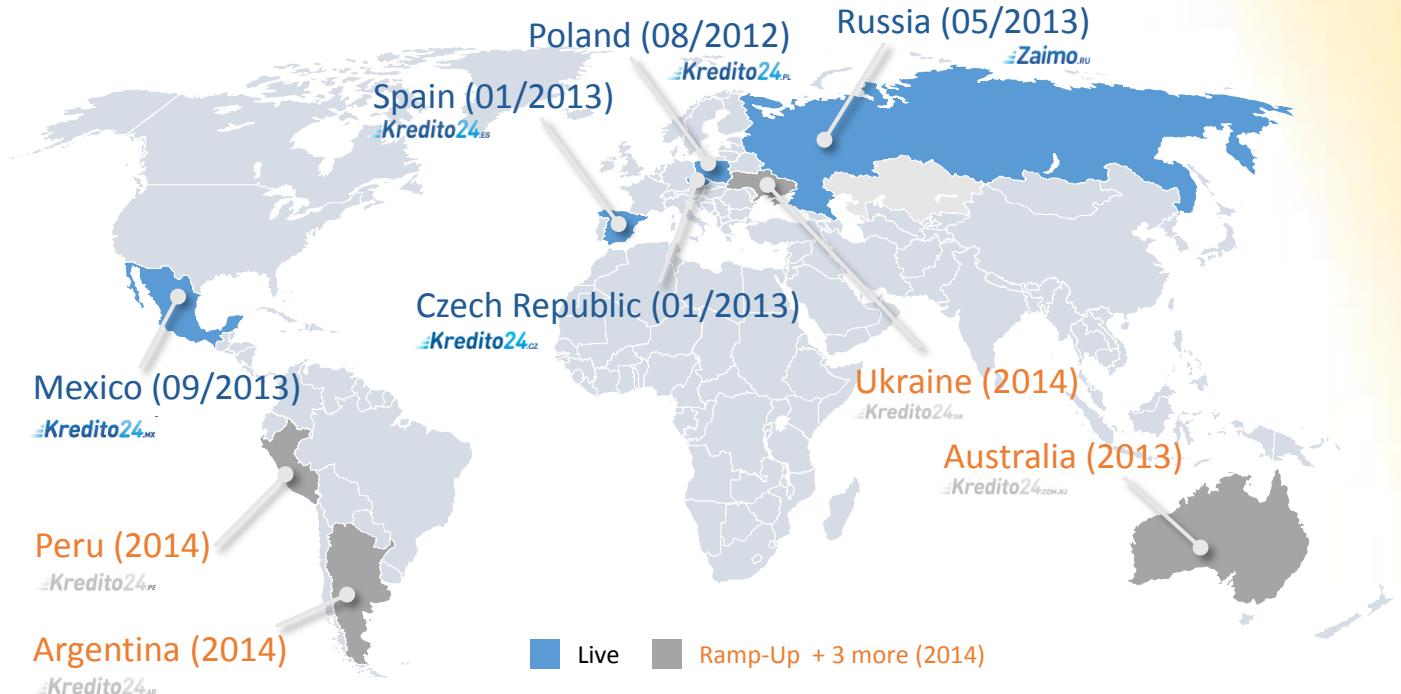
## *Profitable, easy to scale operations across the globe*

### Highly profitable operations

- All live countries are profitable on a unit economics basis
- No local presence required (only legal entity + virtual office)
- Automated operations
- Fast, efficient expansion (~ 4 months, 15k per new country)
- Emerging markets show highest demand, margins and market potential

### Defendable USP

- Technology & data allows to launch countries with virtually no competition at all (for lack of credit bureaus)
- High degree of proprietary tech

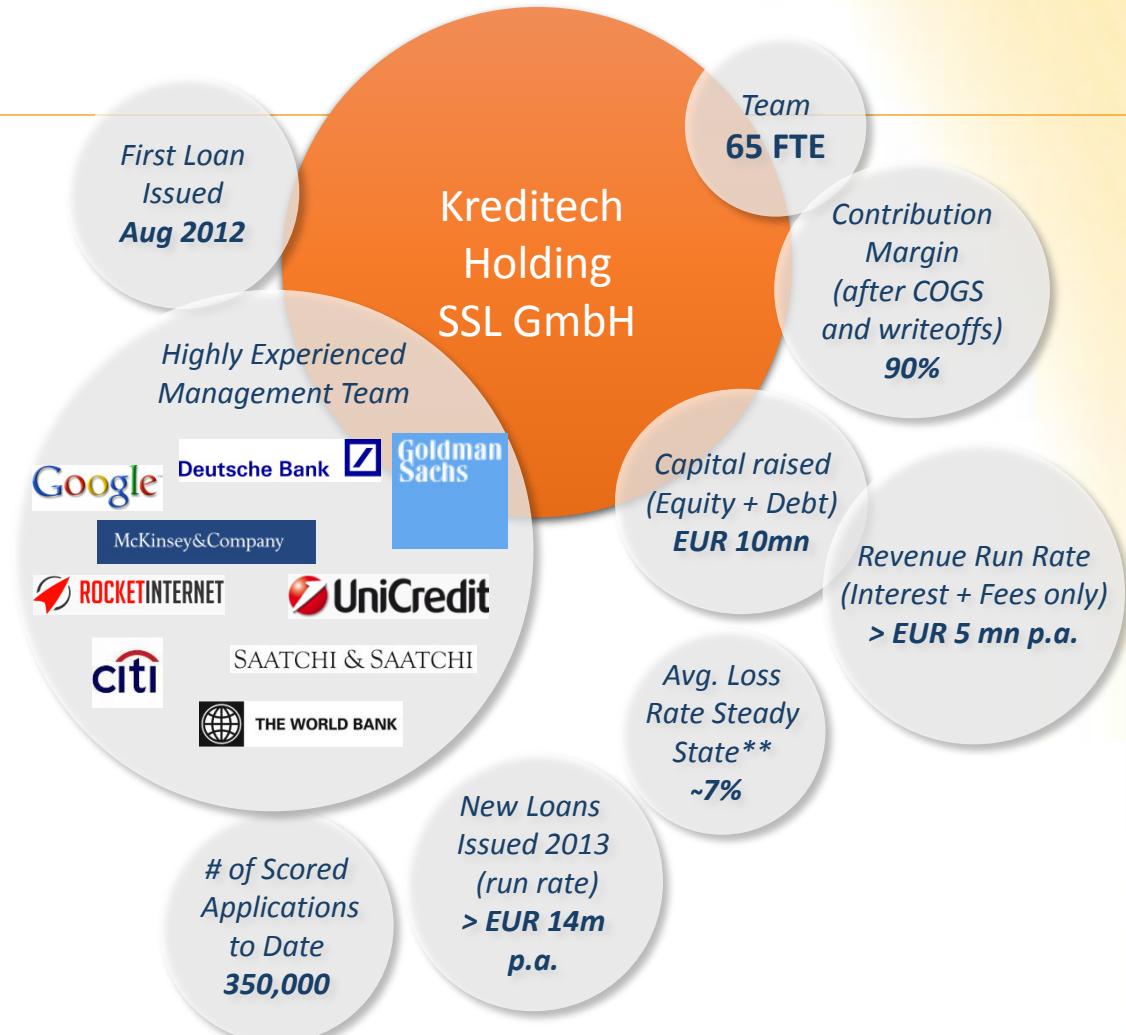


*Short-term loans generate rapidly growing cash flows and serve as data basis for constantly re-iterated scoring technology, which will be gradually rolled out to longer durations*

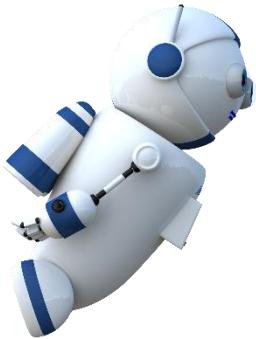
# Kreditech Key Facts

As per November 2013

- 1,200% Growth in 2013 in revenue
- Profitable in core market Poland
- 20 mn USD issued loans in 2013
- Default rate less 10% (group wide)
- Over 250,000 applications till 10/2013
- >1,000 applications / day
- Company Valuation > 70mn EUR
- 10mn EUR raised from Samwers, Silicon Valley, Team Europe, Heiko Hubertz, Michael Brehm, Stefan Glaenzer, etc.



# Contact



Website: [www.kreditech.com](http://www.kreditech.com)

Twitter: @kreditech

Facebook: [www.facebook.com/kreditech](http://www.facebook.com/kreditech)



**Sebastian Diemer**

Founder & CEO of Kreditech

eMail: [sebastian@kreditech.com](mailto:sebastian@kreditech.com)

LinkedIn: [www.linkedin.com/in/sebastiandiemer](http://www.linkedin.com/in/sebastiandiemer)



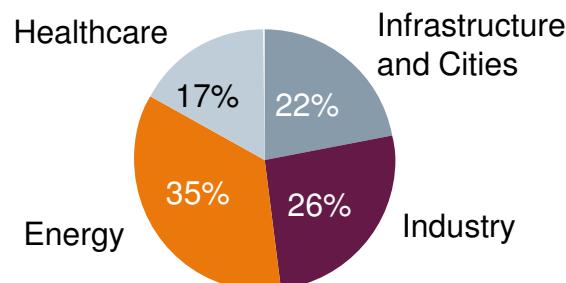
Köln – 26. November 2013

# Sensordaten – Big Data im industriellen Umfeld bei Siemens

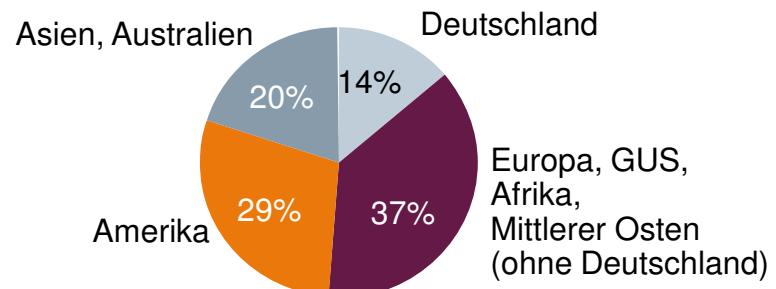
Gerhard Kreß – Siemens AG

# Die Siemens AG ist ein globales Unternehmen mit Fokus auf industrielle Märkte

Umsatz nach Sektoren

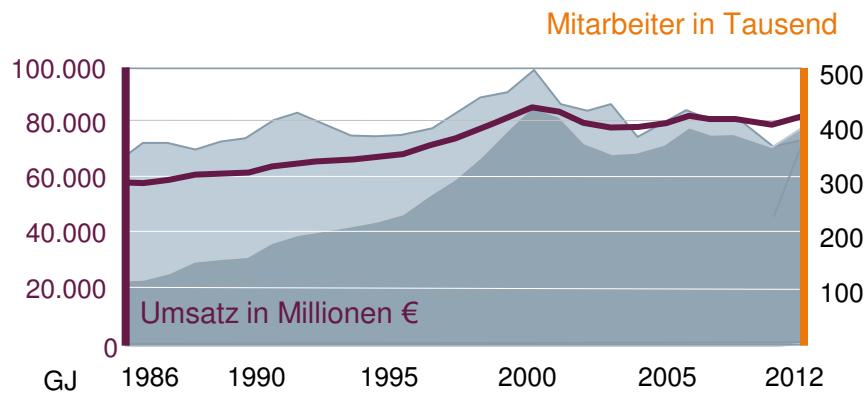


Umsatz nach Regionen



Nach Sitz des Kunden

Entwicklung Umsatz und Mitarbeiter



Vorjahresvergleich fortgeführte Aktivitäten

in Millionen €; fortgeführte Aktivitäten (GJ 2012)	GJ 2011	GJ 2012
<b>Auftragseingang</b>	85.166	76.913
<b>Umsatz</b>	73.275	78.296
<b>Ergebnis</b>	7.376	5.184
<b>Free Cash Flow</b>	5.918	4.790
<b>Mitarbeiter</b>	359.000	370.000

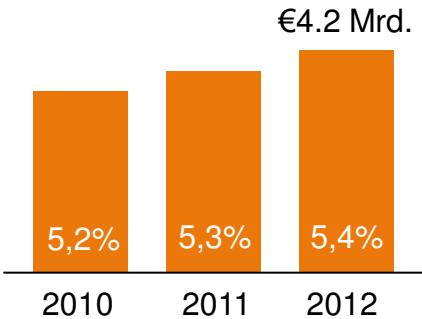
Basis für Zahlen: Jahresberichte des Unternehmens

# Innovationskraft gehört zum Kern von Siemens

Mitarbeiter in Forschung und Entwicklung

- 29.500 Mitarbeiter weltweit
- Rund 17.500 Software-Ingenieure
- Rund 190 FuE-Standorte in 30 Ländern

FuE-Ausgaben in % vom Umsatz



Position bei Patenten (2012)

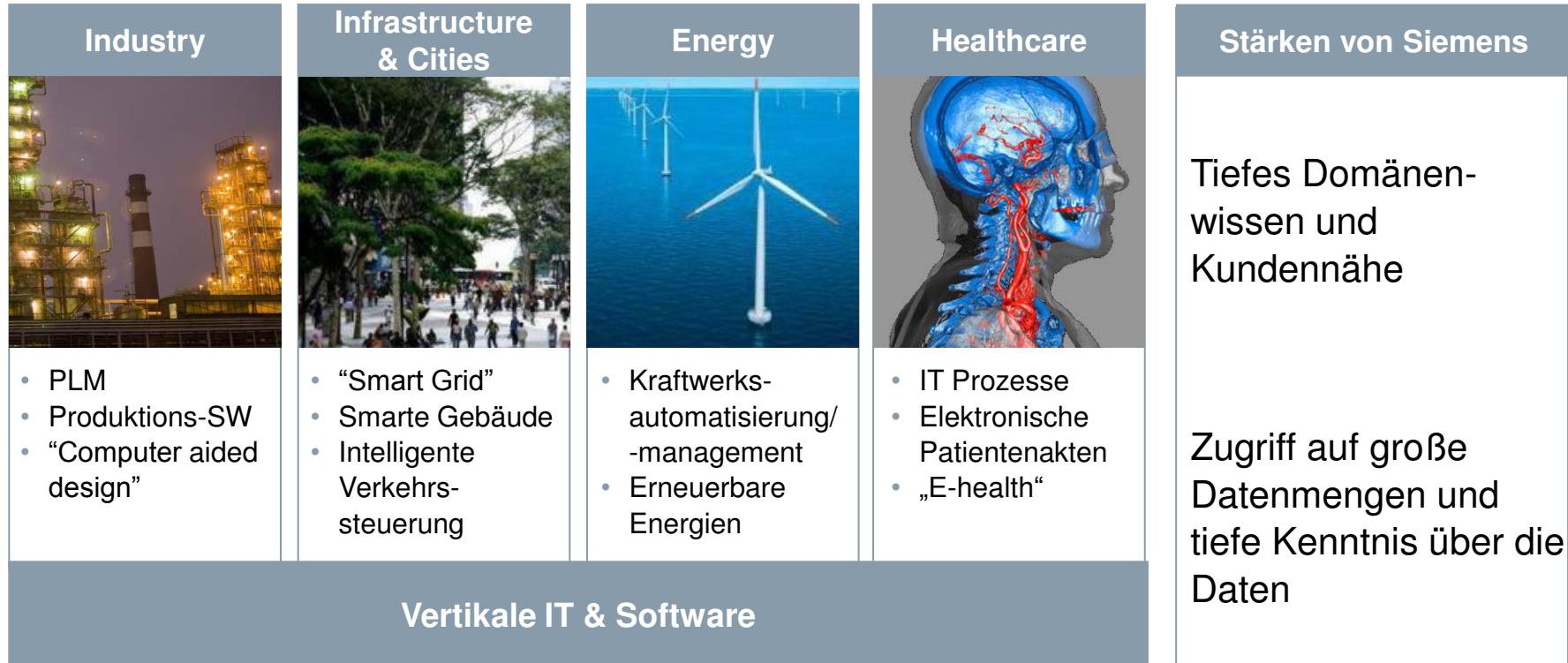
- |              |        |
|--------------|--------|
| Deutschland: | Nr. 3  |
| Europa:      | Nr. 2  |
| USA:         | Nr. 10 |

- 57.300 erteilte Patente insgesamt
- Davon 20.200 „grüne Patente“ zur Absicherung unseres Umweltportfolios
- 9.000 Erfindungen im GJ 2012

Kooperationen

- Pro Jahr über 1.000 neue Partnerschaften mit Universitäten und Forschungseinrichtungen

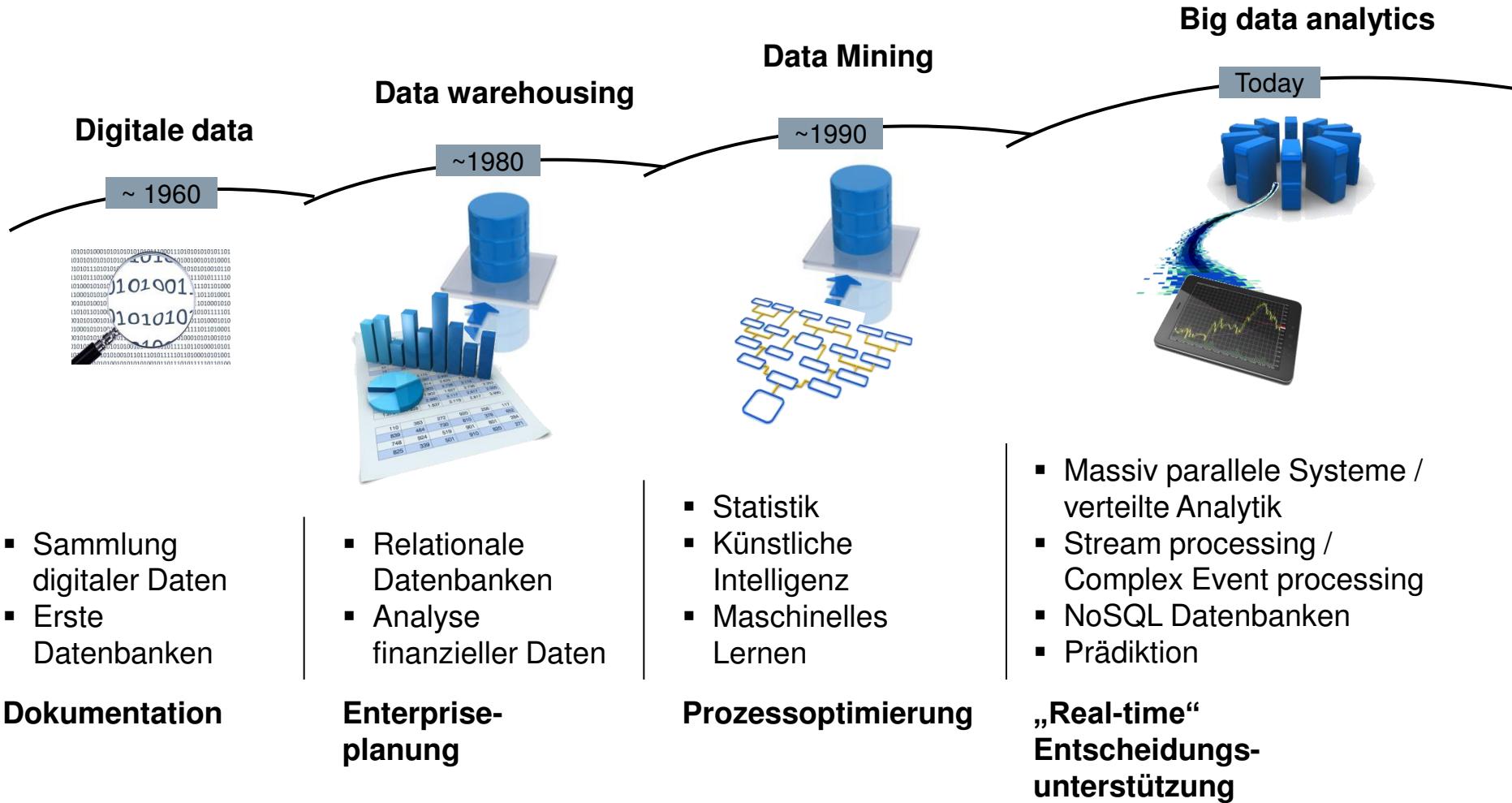
# Siemens fokussiert sich auf „Vertikale IT“ – eine Kombination von Domänenwissen und Technologie



**Horizontale IT**  
(Infrastruktur, Plattformen und Dienstleistungen)

PLM: Product lifecycle management

# Big Data – Eine neue Ära der Daten beginnt gerade



# Industrielle Daten werden in Zukunft einen Hauptfokus von “Smart Data” darstellen

## Heute steht „Social Media“ im Fokus



Facebook wächst um  
**250 Millionen**  
Fotos am Tag

Social Media



Mobile Geräte

## In der Zukunft geht es mehr um Sensordaten



Eine Ölplattform liefert  
**25 Tausend**  
Datenpunkte je Sekunde



Auslesen der „Smart Meters“  
alle 15 Minuten erzeugt  
**3,000 mal mehr**  
Daten

Geophysische  
Exploration

Medizinische  
Bildgebung

Smart  
Grids

Wie können wir aus diesen  
**Daten**  
**Informationen**  
gewinnen und  
**Handlungen**  
daraus ableiten?

# “Smart Data” wird radikale Veränderungen in industriellen Prozessen ermöglichen

## Volumen

Daten **komplexer Systeme** – nicht nur Komponenten

## Geschwindigkeit

Beantwortung von Fragen während sie noch **relevant** sind

## Vielfalt

Einbettung in einen **situativen Kontext**

- Optimierung des Verhaltens komplexer Systeme
- Entscheidungen in operativen Prozessen in Echtzeit
- Automatisierung auf Ebene komplexer Systeme

Kernkompetenz:  
**Smart Data**

# Beispiel: Data Management und Real Time Monitoring bei Gasturbinen



## Module

Online-Daten: ca. 5,000 Datenpunkte / s

Real-time Datenanalyse (1,000 Neuronale Modelle)

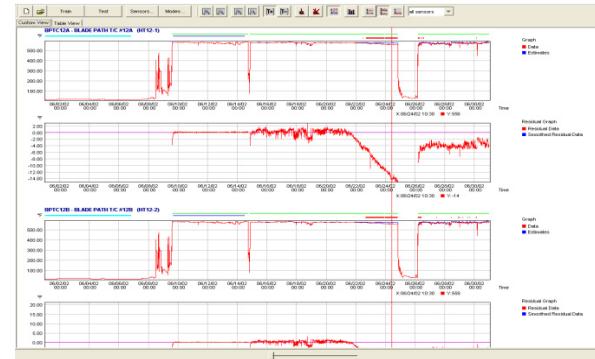
Databank: Eingabedaten und Modellergebnisse

Komplette Datenanalyse und selbstlernende Optimierung

## Vorteile

- Verbessertes Hochfahren der Gasturbine mit verringerten Vibrationen (und geringerem Verschleiß)
- Reduzierte NOx Emissionen
- Verbesserte Effizienz der Gasturbine
- Unterstützung für den Entwicklungsprozess

# Beispiel: “Plant Monitor” entdeckt schleichende Materialermüdung und warnt vor Ausfällen



Plattform, die sich anbahnende Probleme vorhersagt durch:

- Lernen von historischen Daten und Trends
- Zusammenführen mit Regeln und menschlichem Wissen

# Beispiel: Hochgeschwindigkeitszüge - ICE, Velaro

## Vorteile

Durch Outsourcing aller Instandhaltungsmaßnahmen und der vollständigen Verantwortung kann sich der Kunde auf sein Kerngeschäft konzentrieren

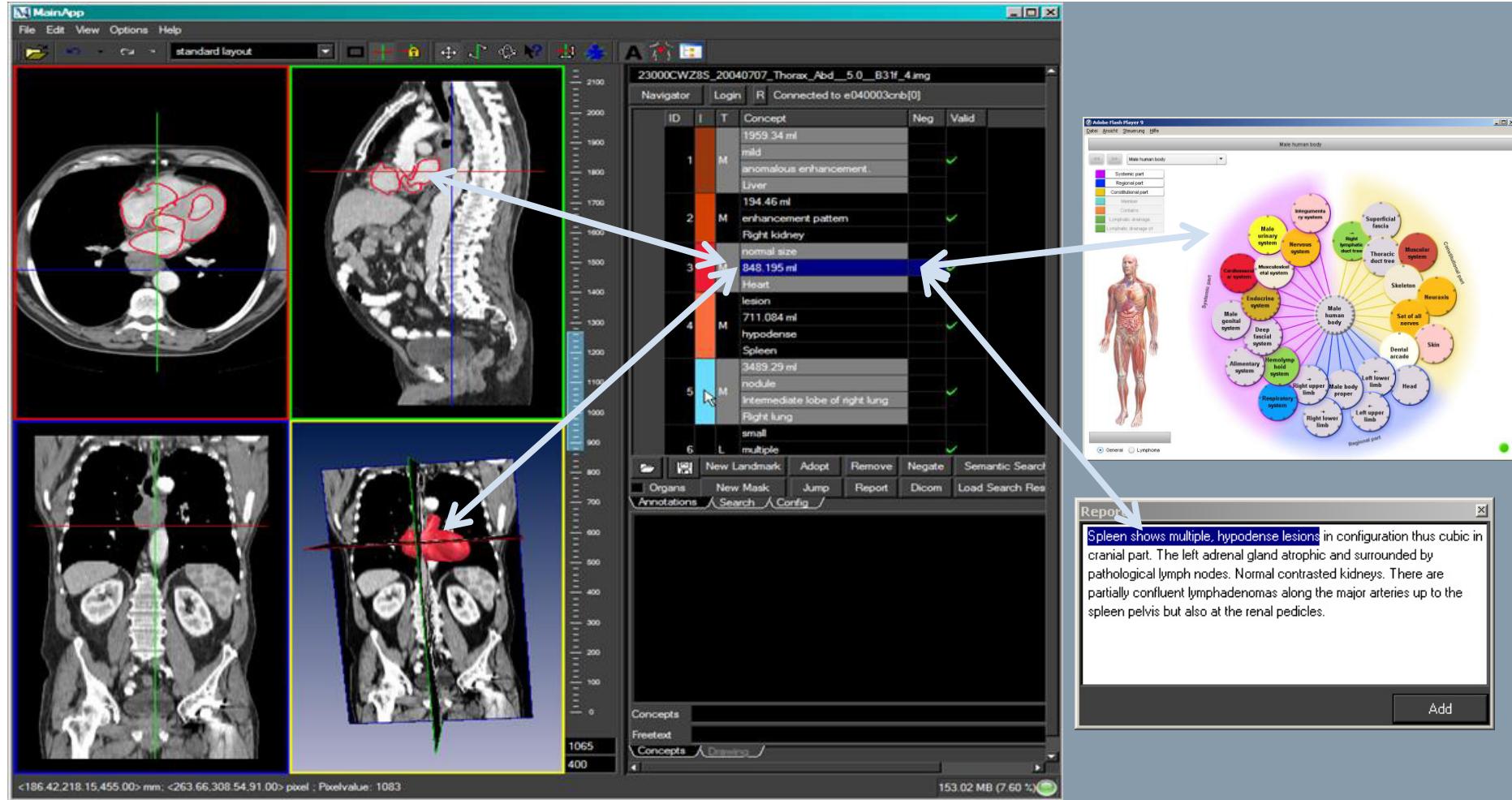


## Ansatz

Zugriff auf den gesamten Bestand aktueller und historischer Betriebsdaten

Einbindung interner und externer Daten zur Produktüberwachung, Wartung und Produktoptimierung

# Beispiel: Semantische Informationen verbinden Texte und Bilder



Verlinkung von Informationen in Texten und Bildern



# Im Takt des Motors - SAP HANA im Prüffeld bei der Mercedes-AMG GmbH

Dirk Zeller, Mercedes-AMG & Dr. Hagen Radowski, MHP  
Handelsblatt Jahrestagung Strategisches IT Management, Januar 2014



Mercedes-Benz

„Driving Performance“ beschreibt die zwei Kernelemente unserer Marke: erstens die Kraftentfaltung, das Handling und den Sound – also die Elemente, die das einzigartige AMG Fahrerlebnis ausmachen; und zweitens unseren Anspruch, der Innovator im High-Performance Segment zu sein.

Der E 63 AMG leistet bis zu 430 kW (585 PS) und 800 Newtonmeter Drehmoment.





DTM-Rennwagen AMG Mercedes 190 E 2.3-16 von 1988.

Jeder braucht etwas, das ihn antreibt. Bei uns ist das die Leidenschaft für Ingenieurskunst und Motorsport.

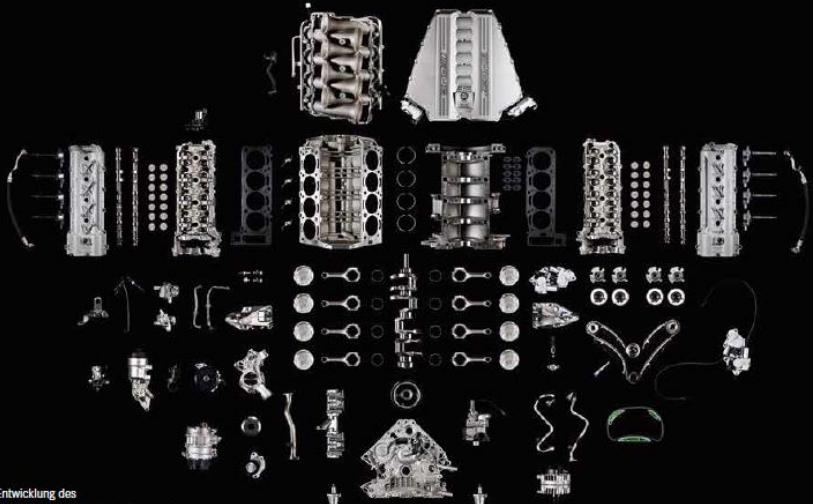
Für die Performance-Marke AMG ist es besonders wichtig, im Motorsport zu bestehen. Denn hier zeigt sich, ob der Transfer von Ingenieurleistungen aus dem Labor in den Grenzbereich funktioniert. Deshalb gehören Motorsport und Ingenieurskunst bei AMG immer zusammen - zum Beispiel bei den Deutschen Tourenwagen Masters.



Geniale Ingenieurskunst  
ist die Summe vieler  
kleiner Verbesserungen.

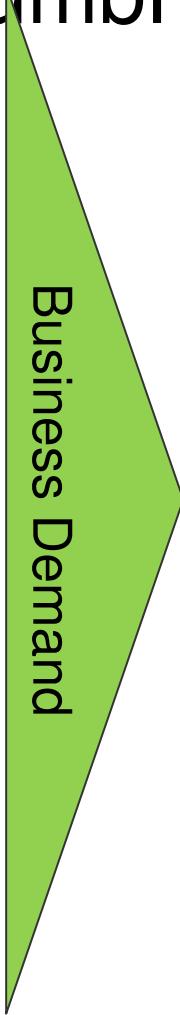
1967: Pleuel des AMG 300 SEL 6.8,  
auch bekannt als „Legende von Spa“.

Von der Liebe zum Detail bis zum Blick fürs große Ganze: In den Anfängen beschäftigte sich AMG hauptsächlich mit der Frage: Welche Teile kann ich verbessern, um ein Auto schneller zu machen? Diese Leidenschaft führte mehr und mehr dazu, dass wir Fahrzeuge ganzheitlich betrachten. Unser aktuellstes Beispiel ist der SLS AMG, der komplett in Affalterbach entwickelt wurde.



2010: Entwicklung des  
M159 AMG 6,3-Liter-V8-Frontmittelmotors.

# In Memory DB und Echtzeit-Analyse als strategischer Ansatz bei der Mercedes-AMG GmbH



- Höhere Effizienz und Zeitgewinn bei der Analyse großer Datenmengen
- Verbesserung der Reaktionsgeschwindigkeit
- Einfache Erschließung komplexer Zusammenhänge
- Automatisierte Nutzung bereits vorhandener Erkenntnisse

# In Memory DB und Echtzeit-Analyse als strategischer Ansatz bei der Mercedes-AMG GmbH



# In Memory DB und Echtzeit-Analyse als strategischer Ansatz bei der Mercedes-AMG GmbH

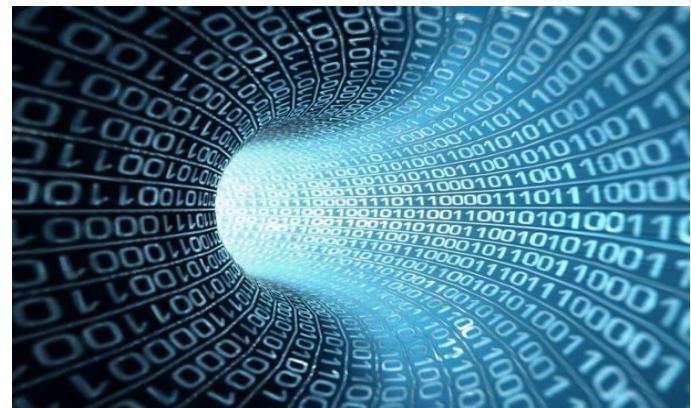


# Innovationsansatz

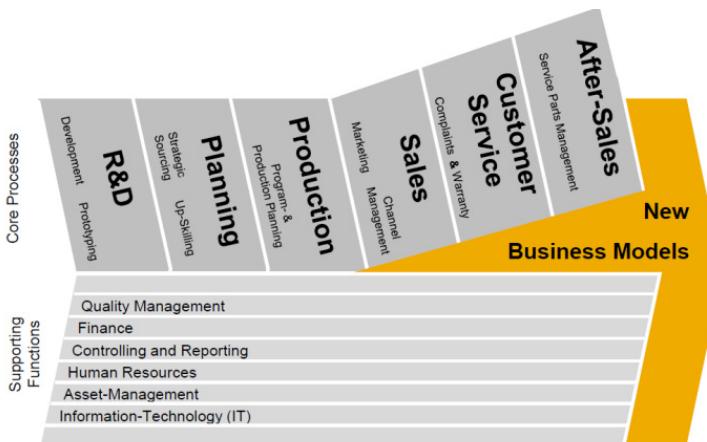
## Realtime & Predictive

## Big Data

OPERATIONS | HR | ACCOUNTING | DISTRIBUTION | MARKETING



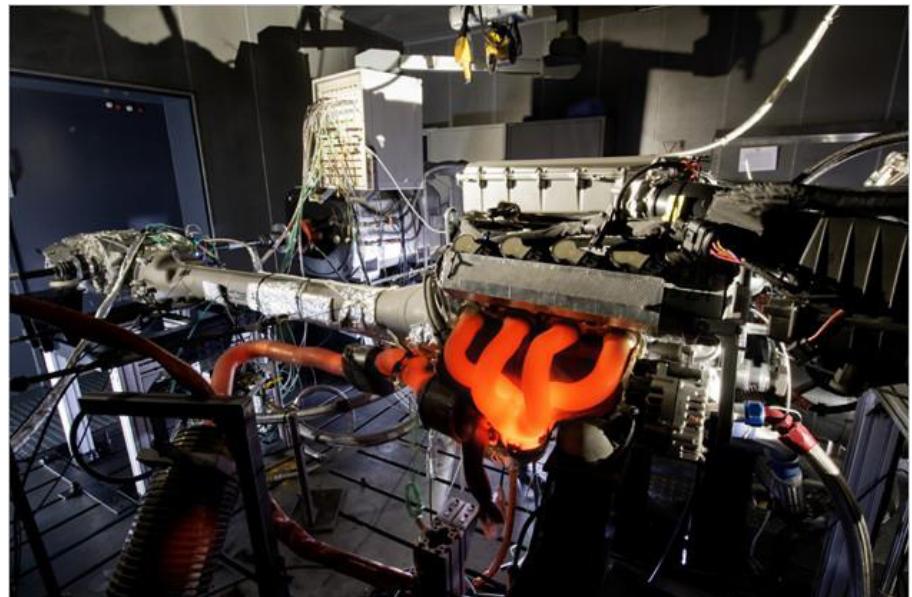
## Kernprozess



# Qualitätssicherung in Echtzeit

## Echtzeit-Auswertung von Messdaten aus Gütesicherungsläufen von High Performance AMG-Motoren auf der Datenbanktechnologie SAP HANA

- Unterstützt den Kernprozess Motorenentwicklung/-erprobung
- Big Data Computing / Echtzeitanalyse
- Nutzung von predictive Funktionen zur Effizienzsteigerung

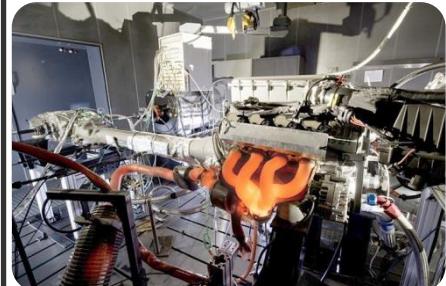


# Das Resultat: Geschäftsprozessinnovation

- **Eine durchgängige Echtzeit-Plattform für Forschung, Entwicklung und Produktion**
  - Vollumfängliche und flexible Echtzeit-Erfassung aller polytechnischen **Motorendaten**: von den Prüfständen über die Teststrecke bis zum Dauerläufertest auf der Straße
  - Echtzeit-Gütesicherung für die Produktion und Echtzeit-Erprobung

## Phase 1

Motor auf dem  
Prüfstand



## Phase 2

Entwicklungsfahrzeu  
ge auf Teststrecken



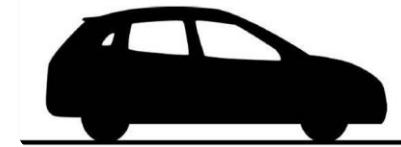
## Phase 3

Dauerlauferprobung,  
Reifegradabsicherun  
g



## Phase 4

Kundenfahrzeuge



— Entwicklung, Erprobung und  
Gütesicherung

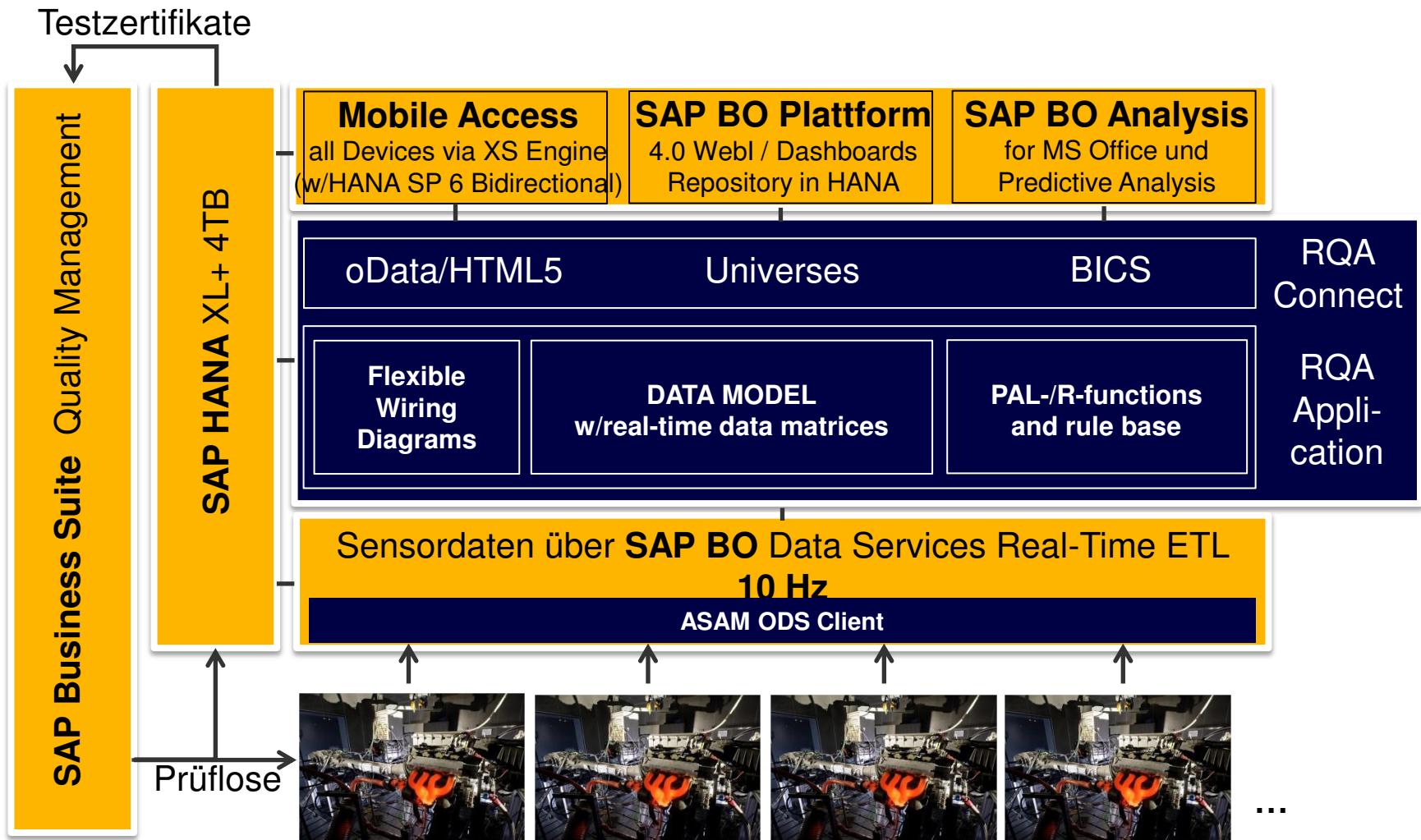
→ SOP • Car IT →

# Die AMG Real-Time Quality Assurance Lösung (RQA)

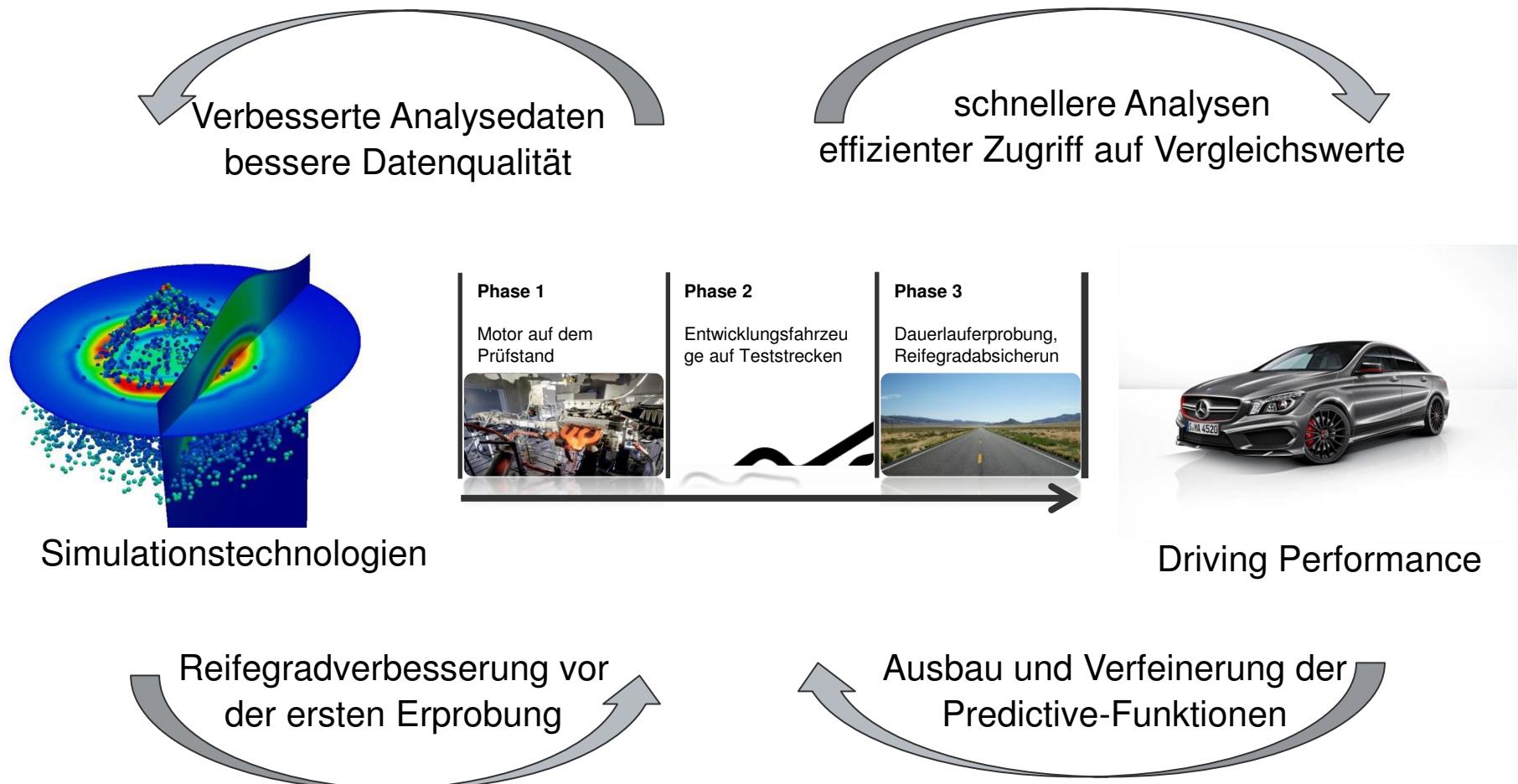
- RQA ist eine leistungsfähige, skalierbare Plattform für die Optimierung von Primärprozessen mit polytechnischen Daten als „Big Data“
- Basiert auf dem internationalen Standard ASAM/ODS, ist im Matrizenkern flexibel konfigurierbar und passt so auf andere „Big Data“ Bereiche wie Windkanal, Crashtests oder Virtual Engineering sowie alle Prüfstände (Getriebe, Bremsen ...)
- Kombiniert drei Innovationen: HANA, Mobile, Predictive Analysis.  
Kann als Stand-Alone Lösung an SAP ERP angeschlossen werden
- Reduziert z.B. die Laufzeit bei Auffälligkeiten im Prüffeld von 50 Min.  
(bisher Standard) auf wenige Minuten (Neu) durch Sofortabbruch in Echtzeit



# RQA technische Architektur „All-in-One“



# Schnelle Analysefunktionen wirken auf vor- und nachgelagerte Prozesse





# Big Data & Business Analytics

Capgemini @ Big Data Europe, Zürich

Jürgen Düvel



People matter, results count.

# Das Phänomen Big Data mystifiziert die Informationswelt



Lichtfeld-  
fotografie

# Einige Annahmen können dem Potential von Big Data im Wege stehen

1

Mit einer Technologie ist Big Data „gelöst“.

Nein, die Technologie ist der „Enabler“, für den Umgang und die Verwendung von Big Data gem einer Informationsstrategie.

2

Mit Big Data entsteht eine komplett neue analytische Lösung.

Nein, es geht darum, bestehend BI Landschaften zielgerichtet – i.S. von Big Data – zu ergänzen und zu verbessern.

3

Big Data ist durch die 3Vs hinreichend erklärt.

Nein, die Aufzählung von Eigenschaften der Daten reicht nicht aus. Die Anwendung von Big Data ist entscheidend.

# Die aktuelle Technologie ermöglicht es uns erst, über die Anwendung von Big Data realistisch nachzudenken

1

Mit einer Technologie ist Big Data „gelöst“.

2

Mit Big Data entsteht eine komplett neue analytische Lösung.

## Technologische Aspekte

- Die **Vermeidung** einer zu grossen **Technologievielfalt** trotz verschiedener Formate, Typen und Quellen von Big Data
- Die **Weiterentwicklung von bestehenden BI und DWH Lösungen**, um für die agile Anwendung von Big Data bereit zu sein
- Die **automatisierte und generische Verarbeitung** in der Aufbereitung und der Analyse von Big Data für (complex event processing).

# Die Kernaufgabe der IT ist es, die Integration, die Stabilität und die Flexibilität ihrer BI Systeme beizubehalten und auszuweiten.

## Analytische Applikationen

Analyse &  
Reporting

Dashboards &  
Scorecards

Mining &  
Profiling

Segmentierung &  
aCRM

Planung &  
Budgetierung

Prognose &  
Simulation

Content  
Analyse

## Business Analytics Integration

Synchronisieren, Korrelieren, Filtern, Klassifizieren, Kommentieren

Data Warehouse

Streaming

Big Data Technologie

## Datenintegration

ERP, 3<sup>rd</sup> Party  
(Kerndaten)

Informationsbroker  
(Feeds)

Social Media  
(Platforms, Tweets)

Klassische  
Massendaten  
(Protokolle)

# Big Data muss organisatorisch, im Programmmanagement und konform zu den Regulatorien auf soliden Füßen stehen.

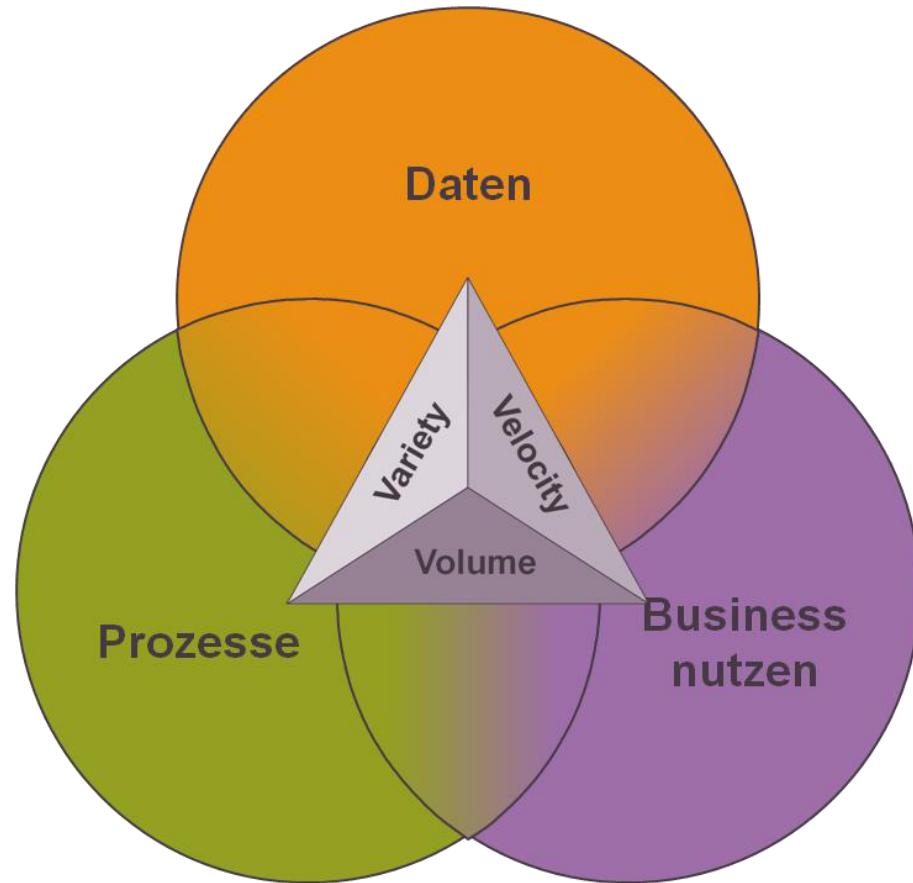
3

Big Data ist durch die 3Vs hinreichend erklärt.

## Organisatorische Aspekte

- Die **Ausrichtung** der Big Data Aktivitäten an der **Informationsstrategie** und dem **Businessnutzen**.
- Die **Praktikabilität** der **Datenmanagement** Disziplinen Governance, Stewardship sowie Meta Data und Data Quality
- Die **Gewährleistung** der Balance zwischen der Business Analytics mit Big Data und der **Datensicherheit**, dem **Datenschutz** und der **Unternehmensreputation**.

# Die Eigenschaften von Big Data dimensionieren die Komponenten von Big Data



- Umkehrung des Informationsprinzips von inside-out nach outside-in
  - Informationen ermöglichen eine Evolution in den Geschäftsprozessen
- 
- Für die Halbwertszeit von Big Data benötigt es reife und stabile Prozesse im Datenmanagement.
  - Big Data ohne Glaubwürdigkeit, Nachvollziehbarkeit und Zweckdienlichkeit ist wertlos.
- 
- Die Bildung eines 360° Information Asset
    - Online Business, Mobile Computing und Social Media
    - Sensoren und RFID Tags
    - Audio, Video

# Der effiziente Umgang mit Big Data basiert auf einer Strategie, einer Methodologie und reifen Lösungsfähigkeiten

## ■ Big Data Prozessmodell



## ■ Zentrale Aspekte für die Entwicklung und Implementierung





# Der Umgang mit Big Data am Beispiel Social Media

# Social Media: Datenlieferant und Aktionsplattform für Big Data in der Kundenkommunikation

Das Datenvolumen der vielfältigen Kundeninteraktion wächst enorm (Juni 2011)



**90 Millionen**

Tweets pro Tag



**24 Stunden**

an Video Uploads bei YouTube  
pro Minute



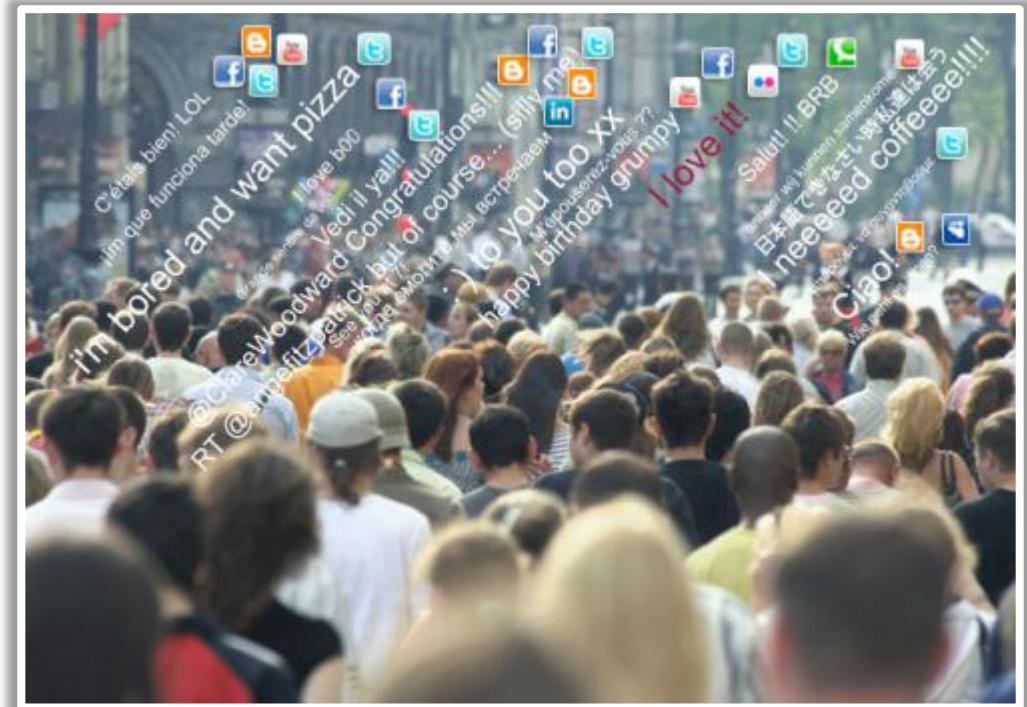
**600 Millionen**

Facebook Anwender weltweit



**24 Stunden**

hat Charlie Sheen gebraucht, um  
1 Milion Followers auf Twitter zu  
bekommen.



Die Kontrolle der Interaktion hat sich auf den Kunden verlagert. Die Kunden definieren wie, wann und wo sie mit Marken, Inhalten & Diensten interagieren.

# Das Vorgehen einer Social-Media-Analyse lässt sich u.a. auf Sensoren-, OnlineShop-, Logging-Analysen adaptieren



Verringerung des Rauschens



Einbringen anderer strukturierter Kunden- & Unternehmensdaten



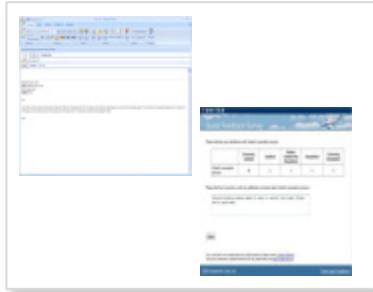
Massnahmen



Social Media Monitoring



Textanalysen



integriertes Wissen



# Kombinieren Sie Social Media Monitoring (SMM) mit Textanalysen und strukturiertem Inhalt

## Schritt 1: Kombination verschiedener Analyseschichten:



 Textanalyse kann die Genauigkeit des SMM stark verbessern, besonders, wenn sie mit anderen wortgetreuen Quellen verwendet wird. Die Kombination strukturierter und unstrukturierter Daten reichert das Wissen an.

# Etablieren Sie eine „Kommandozentrale“, damit die Anwender Entscheidungen nach Mass treffen können

**Schritt 2: Agieren Sie konsequent Informationsgestützt:**



 Technologie wird niemals zu 100 % genau sein. Menschliche Intervention ist gefragt, um Ausgewöhnliches zu erkennen, wenn es an die Oberfläche tritt, um Daten zu korrigieren und um Entscheidungen über Massnahmen zu treffen.

# Die gewünschte Kundenerfahrung muss in Geschäftsregeln mit den benötigten Informationen abgebildet sein

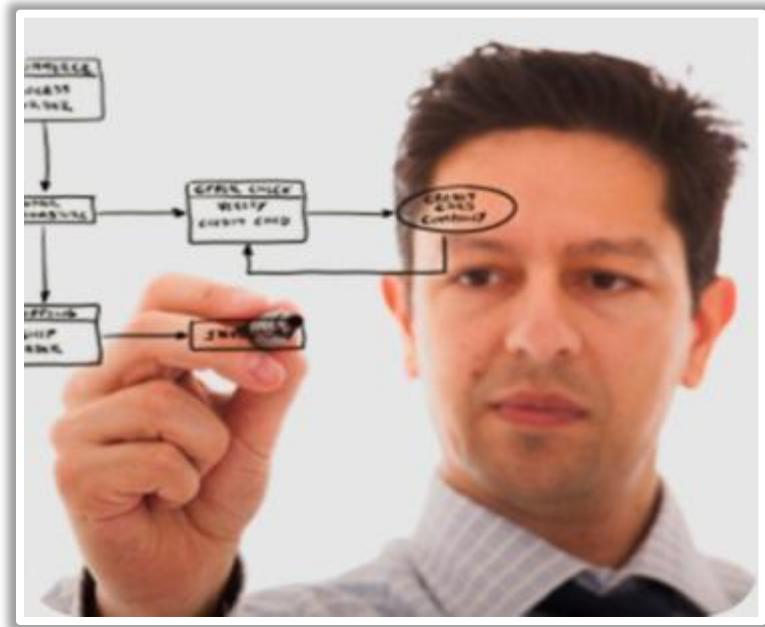
**Schritt 3: Definieren Sie Informationsobjekte und –prozesse mit Geschäftsregeln:**



Irgendwo im Rauschen der sozialen Medien sind konkrete Chancen verborgen, um das Problem eines Kunden zu beheben, um ein Produkt oder eine Dienstleistung zu verbessern oder um eine neue Verkaufschance zu identifizieren.

# Weg von der reinen „Kundenkontaktzentrale“, hin zur handlungsfähigen und proaktiven Betreuung.

**Schritt 4: Verknüpfen Sie Ihr Fallmanagement mit Ihren Geschäftsprozessen:**



 Wenn Kunden twittern und bloggen, haben sie üblicherweise einen Grund dafür. Ein Eins-zu-Eins-Feuerlöschmodell wird sich nicht an das Volumen der sozialen Medien anpassen können - errichten Sie bloss nicht noch eine teure Kontaktzentrale!

# Ein Auswahl der Nutzung von Big Data in Ihrem Unternehmen

## Kunden-service

- Bereitstellung von Dienstleistungen bevor Sie eine Beschwerde erhalten
- Ermöglichung des gegenseitigen Supports des Kunden
- Ständige Multi-Kanal-Erfahrung

## Produkt-entwicklung

- Optimiert das Produkt- und Dienstleistungsportfolio
- Nutzt die Kreativität & das Wissen der Masse

## Unter-nehmens-strategie

- Kundenorientierung
- Trends erkennen
- Verbessert die Medien-Berichterstattung
- Wettbewerbsbeobachtung

## IT Service Agreements

- Umfassendes, reaktionsschnelles Monitoring
- Proaktiver Support auf Basis des Logs
- Anforderungen aus Benutzerverhalten ableiten

## Marketing & PR

- Ermöglicht die Interaktion unter Benutzern zu ihren Bedingungen
- Steigerung der Markenreputation
- Besseres Verständnis der Kundenbedürfnisse

## Vertrieb

- Eventgesteuertes Inbound-Marketing
- Individualisierung des Kundenkontakt
- Transparenz und sofortiges Feedback über Peer-to-Peer Reviews & Empfehlungen

## HR

- Loyalität , Zufriedenheit & Reputation aus Mitarbeiteraussagen
- Finden Sie neue Talente & optimieren Sie die Rekrutierung

## Risiko / Compliance

- Verbesserte Fähigkeit zur schnelleren Identifizierung und Reaktion auf betrügerische Aktivitäten

# Zusammenfassung der Business Analytics mit Big Data

Daten entstehen ausserhalb Ihres Unternehmens. Wertvolle Information kommt outside-in!



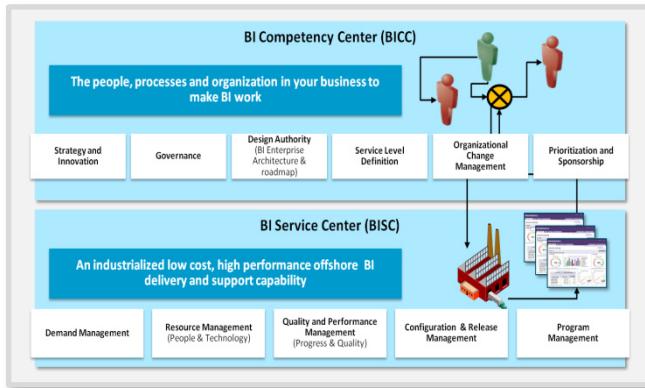
Machen Sie Big Data nutzbar. Überführen Sie ihre Erkenntnisse in Handlungen. Steuern Sie.



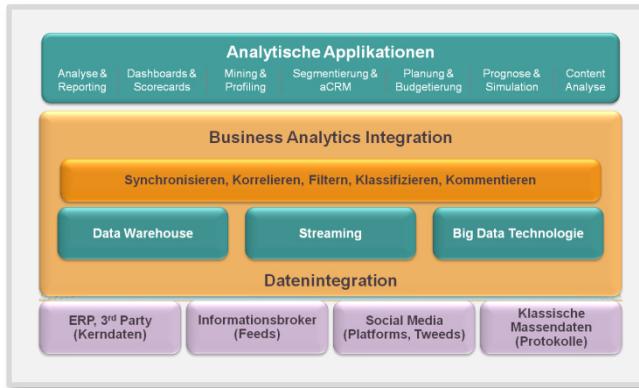
Sie müssen den Hype durchdringen, das Rauschen herausfiltern und fokussieren Sie auf geschäftsrelevante Informationen.

# Capgemini deckt Handlungsfelder von Big Data mit bewährtem Framework für das Business Information Management ab

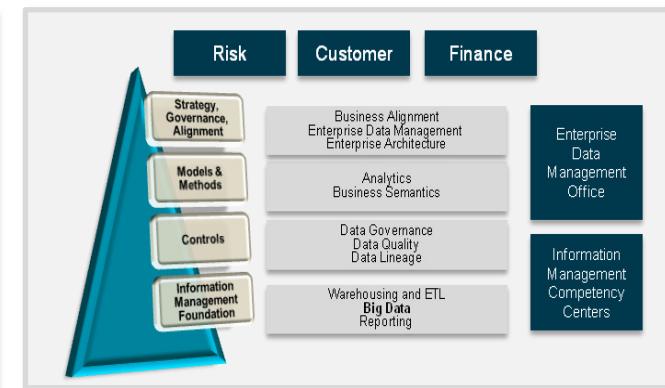
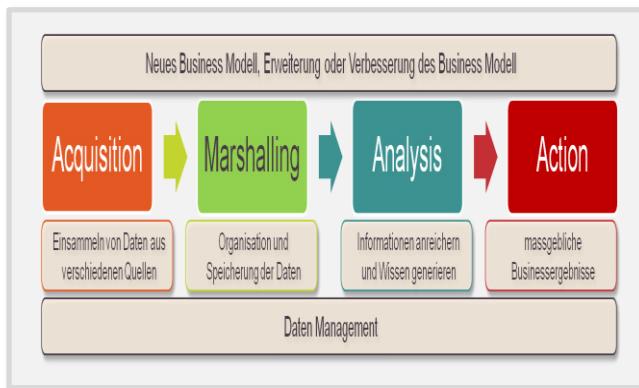
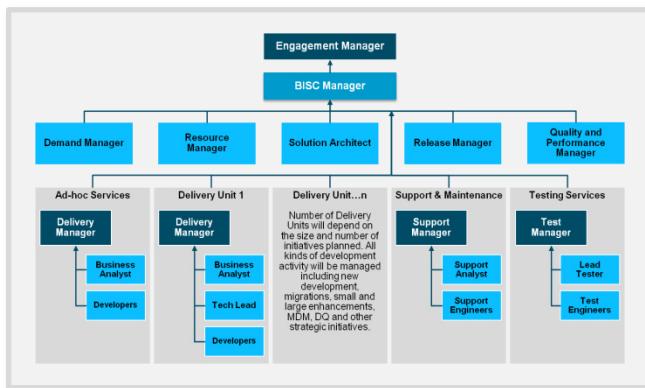
## Governance, BI Service Center, BI Competency Center



## Referenzarchitekturen, Methodologien & Strategien



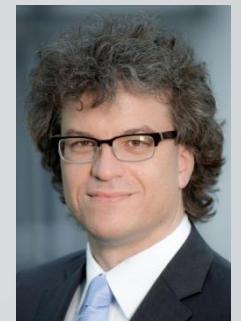
## Technologien und Implementierung



# Vielen Dank für Ihre Aufmerksamkeit !



Please contact me!  
[juergen.duevel@capgemini.com](mailto:juergen.duevel@capgemini.com)  
079 959 00 89



**People matter, results count.**

[www.capgemini.com](http://www.capgemini.com)

# Microsoft is big in BIG DATA

Daniel Weinmann, Stephan Pepersack  
Product Marketing Manager



# A NEW SET OF QUESTIONS



What's the social  
sentiment for my  
brand or products



How do I optimize my  
fleet based on weather  
and traffic patterns?

How do I better  
predict future  
outcomes?

ADVANCED ANALYTICS

# NEW OPPORTUNITIES



## Revenue Growth

Increases ad revenue by processing 3.5 billion events per day



## Businesses Innovation

Measures and ranks online user influence by processing 3 billion signals per day



## Operational Efficiencies

Uses product/ usage analysis and reseller fraud detection for its services

## Massive Volumes

Processes 464 billion rows per quarter, with average query time under 10 secs.

## Cloud Connectivity

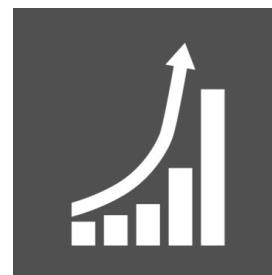
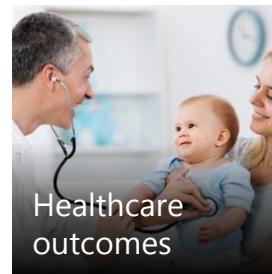
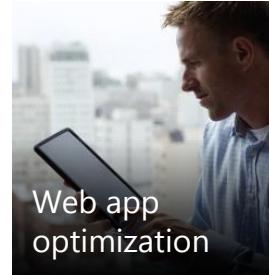
Connects across 15 social networks via the cloud for data and API access

## Real-Time Insight

Improves operational decision making for IT managers and users

# COMMON BIG DATA CUSTOMER SCENARIOS

GAIN COMPETITIVE ADVANTAGE BY MOVING FIRST AND FAST IN YOUR INDUSTRY





# Microsoft brings Big Data to the masses



Relational



Non-Relational

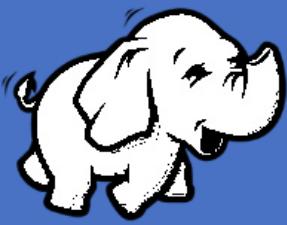


0101010101010101  
1010101010101010  
01010101010101  
101010101010

A visualization of binary code (0s and 1s) arranged in a grid-like pattern, with a translucent 3D cube overlay, representing streaming data.

Streaming

# HADOOP INTEGRATED INTO THE DATA PLATFORM



Non-  
Relational



Windows Server



Windows Azure™

**Microsoft HDInsight Server** for on-premises  
**Windows Azure HDInsight Service** for cloud

Enterprise class security, HA & management  
Seamlessly integrated with Microsoft BI tools  
Windows Simplicity and Manageability  
Provisioned in minutes on Windows Azure

BUILT ON HORTONWORKS DATA PLATFORM (HDP)



manage



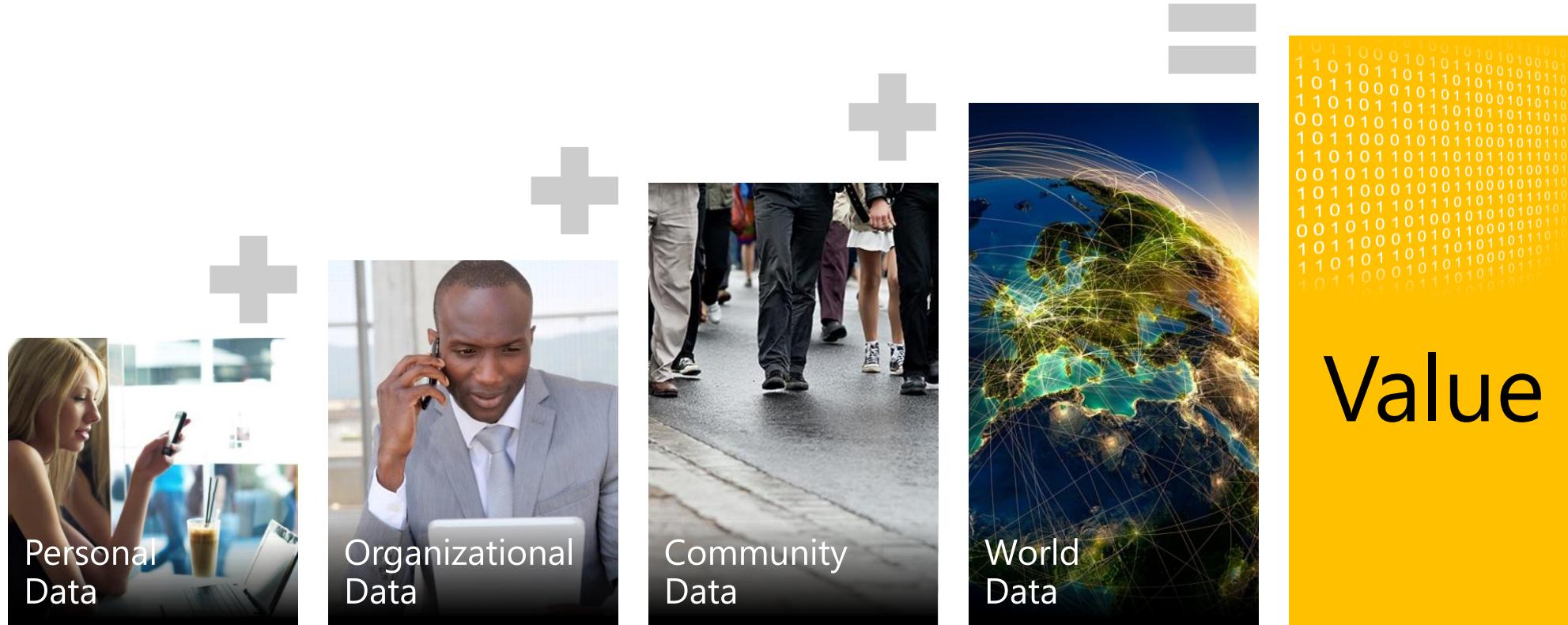
enrich



insight

enrich by connecting to the world's data

# POWER OF COMBINING THE WORLDS DATA





manage



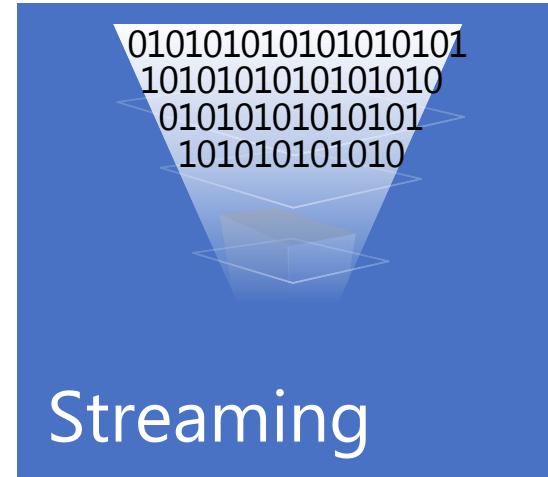
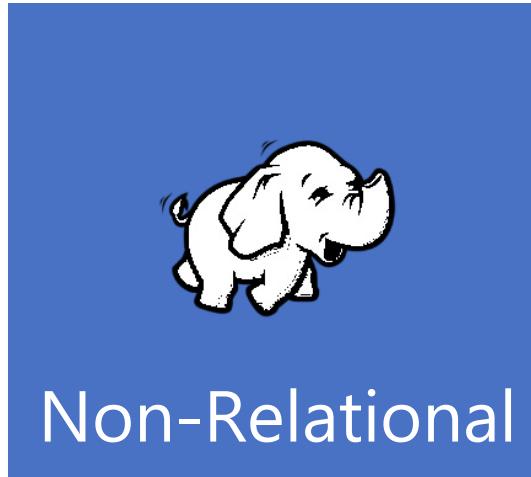
enrich



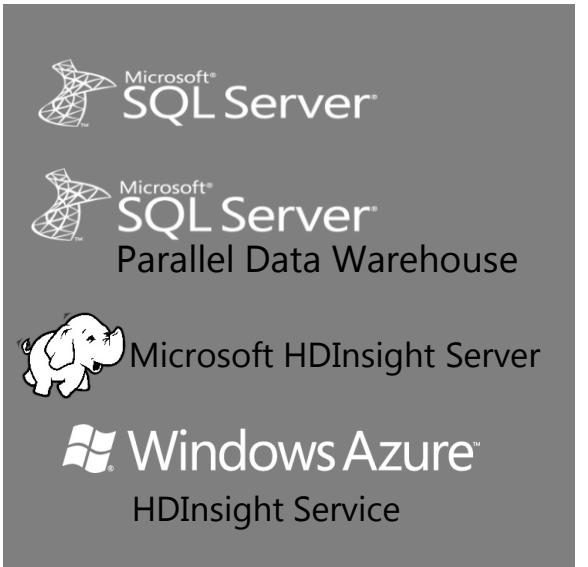
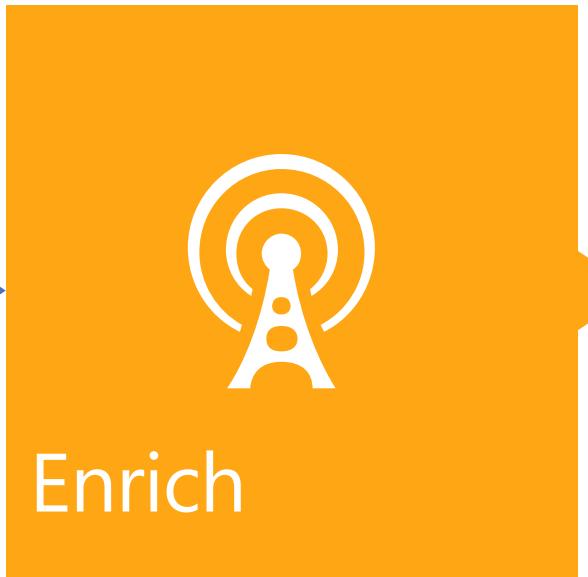
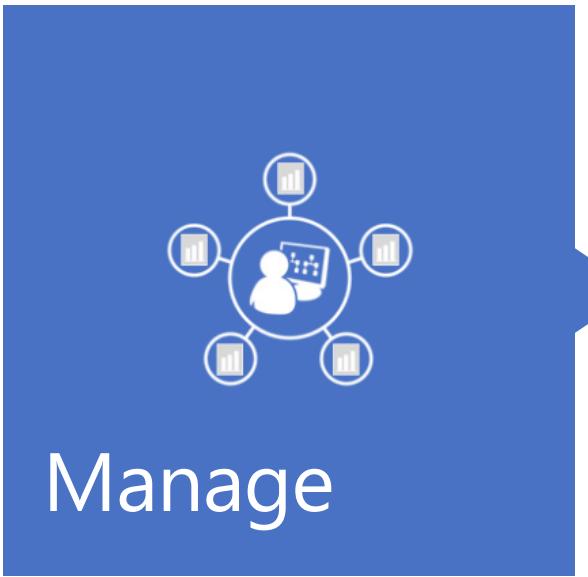
insight

insights for everyone

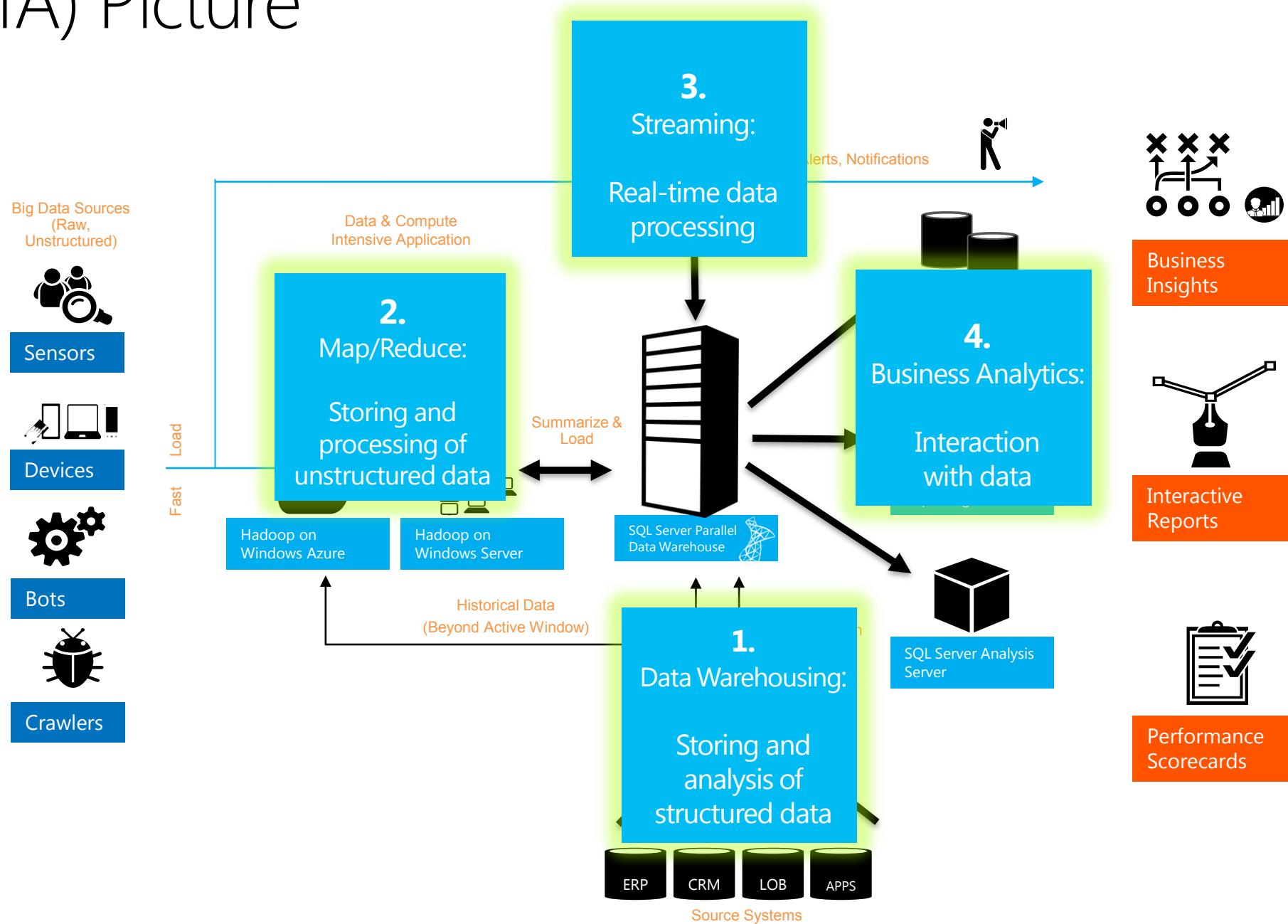
# INSIGHTS ON ANY DATA, ALL USERS, WHEREVER THEY ARE



# MICROSOFT BIG DATA



# Big (DATA) Picture



# Additional Resources

- LEARN MORE
  - Microsoft Big Data Solution: [www.microsoft.de/bigdata](http://www.microsoft.de/bigdata)
  - Windows Azure: [www.windowsazure.de](http://www.windowsazure.de)
- TRY NOW
  - Preview of the Windows Azure HDInsight Service:  
<https://www.hadooponazure.com>
  - Developer CTP of Microsoft HDInsight Server for Windows Server: <http://www.microsoft.com/bigdata>

**Gewinnen Sie ein Microsoft Surface.**

Jetzt hier am Microsoft Stand registrieren.





© 2012 Microsoft Corporation. All rights reserved. Microsoft, Windows, Windows Vista and other product names are or may be registered trademarks and/or trademarks in the U.S. and/or other countries.  
The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation. MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.