

SYMPOSIUM WIRTSCHAFTSINFORMATIK

Neue Entwicklungen in der Wirtschaftsinformatik

Ruhr-Universität Bochum

14. November 2014

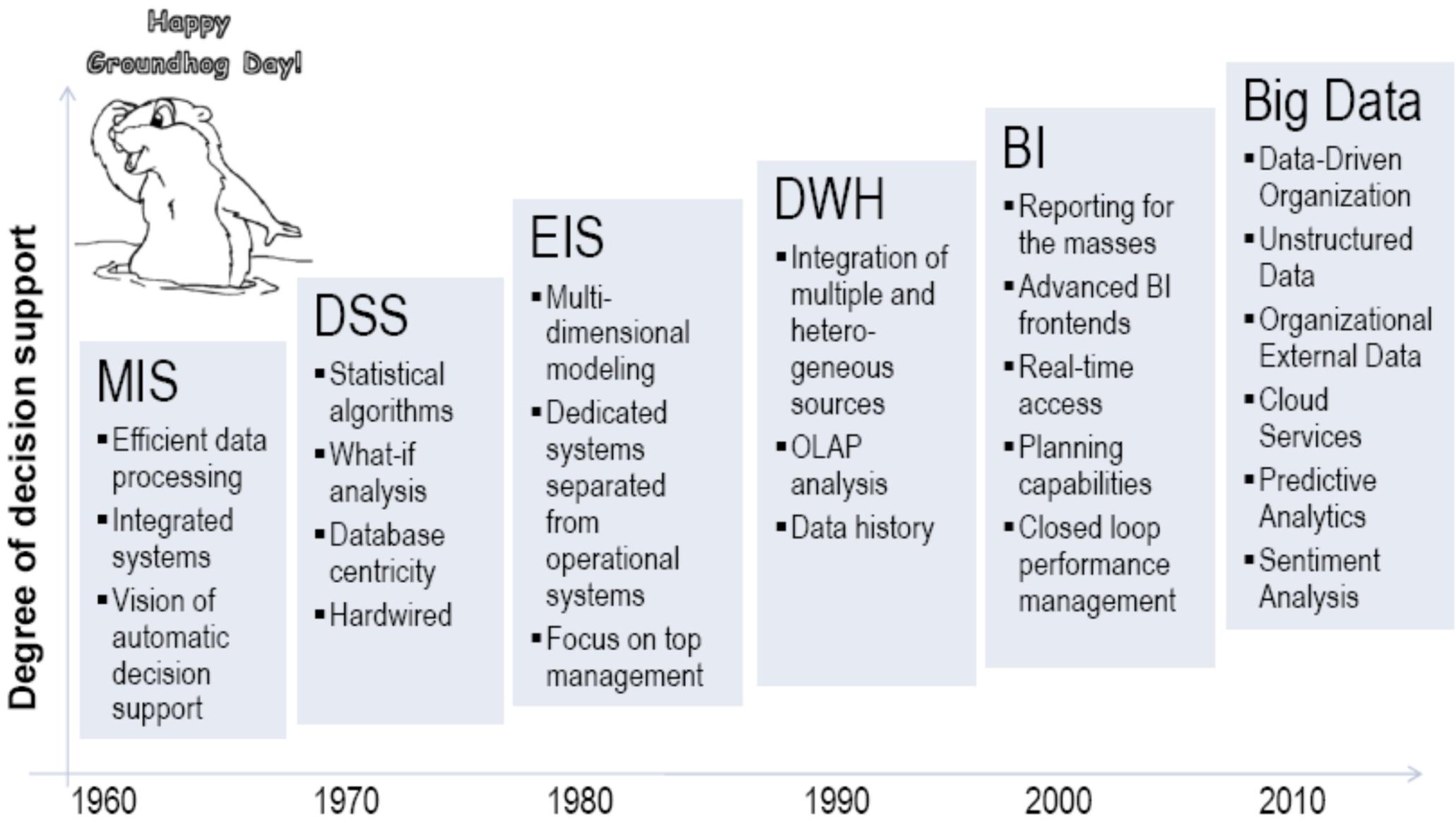
Neue Entwicklungen im Business Intelligence

Univ.-Prof. Dr. Peter Chamoni

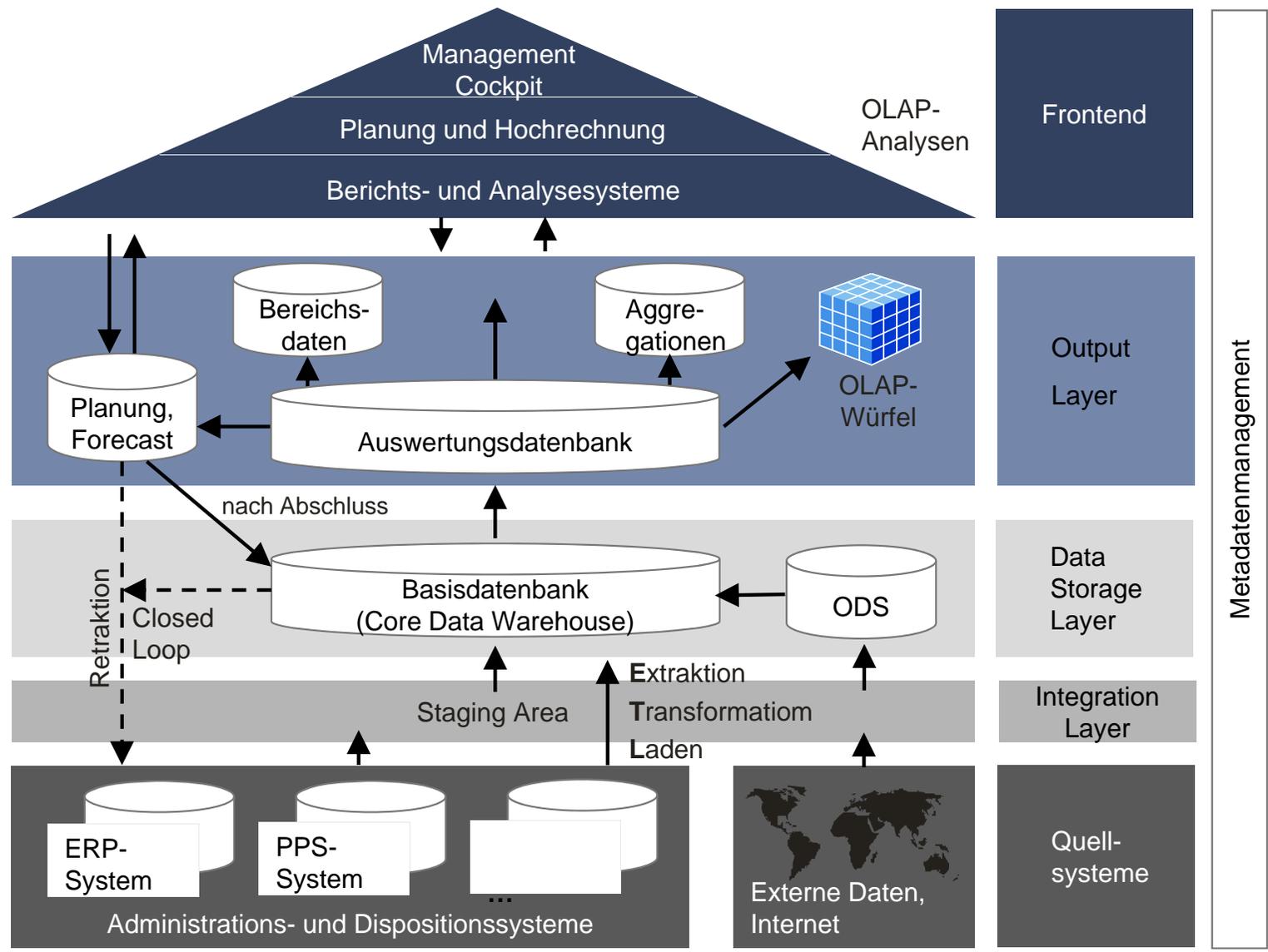
Agenda

- 1 Einleitung
- 2 Big Data
- 3 Empirische Studien
- 4 Markt, Tools und Anwendungen
- 5 Zusammenfassung und Ausblick

1 Evolution der MSS

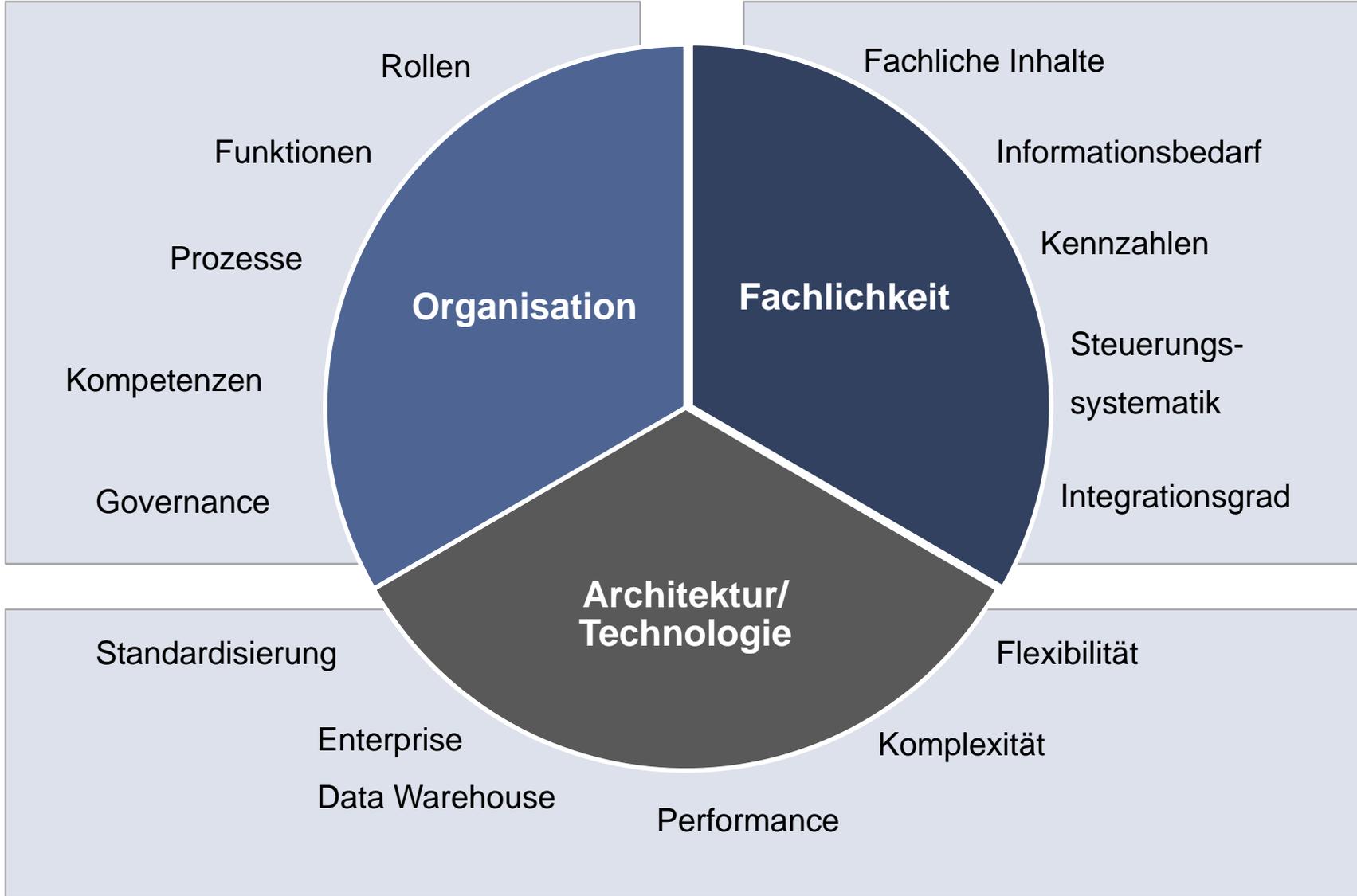


1 Data Warehouse / BI



[nach Gansor et al., 2010]

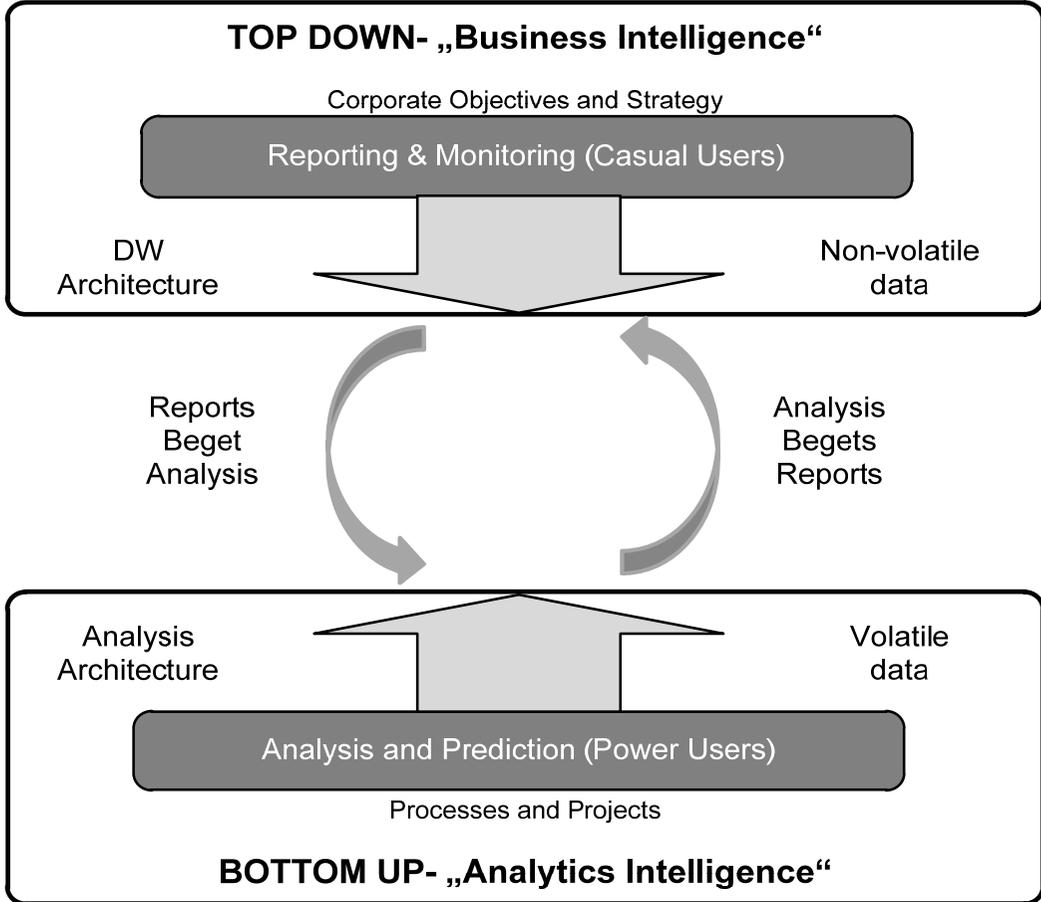
1 BI Maturity



1 Top-down vs. Bottom-up BI

Top-down vs. Bottom-up BI

Top-down and bottom-up BI environments are distinct, but complementary, environments, but most organizations try to shoehorn both into a single architecture.



[Eckerson 2011]

Agenda

- 1 Einleitung
- 2 Big Data
- 3 Empirische Studien
- 4 Markt, Tools und Anwendungen
- 5 Zusammenfassung und Ausblick

BIG DATA

A black and white photograph of a person on a bicycle carrying a massive, precarious stack of cardboard boxes. The stack is tilted and nearly obscures the person, symbolizing the overwhelming volume of big data. The person is wearing a hat and a patterned shirt. The background shows a street with a tree and a building.

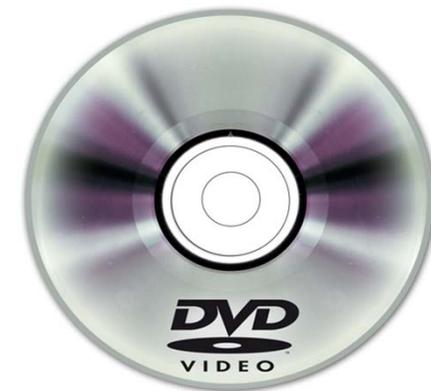
Die Zukunft der Datenanalyse?

2 Anekdote

Die von EMC gesponserte „Digital Universe Study“ beziffert die 2012 erzeugte Datenmenge weltweit auf 2,8 Zetabytes.

Das entspricht 2,8 Billionen Gigabytes und damit...

... 754.000 Jahren Filmen in DVD-Qualität oder aber...



[Digital Universe Study 2012], Zahlen: Eigene Recherche

2 Anekdote

... etwa 560.000.000.000.000.000 (560 Billionen) Seiten Roman.



Aufeinander gestapelt ergäbe das eine Strecke, die 373 mal zur Sonne reichen würde und außerdem...

[Digital Universe Study 2012], Zahlen: Eigene Recherche

2 Anekdote

... müsste zur Produktion dafür ein Wald abgeholzt werden, der vollständig Europa, Nordamerika, Südamerika und Afrika bedecken würde (jedenfalls bis Ende 2012, dann wäre er verbraucht).



[Digital Universe Study 2012], Zahlen: Eigene Recherche

2 Gartner's Hype Cycle

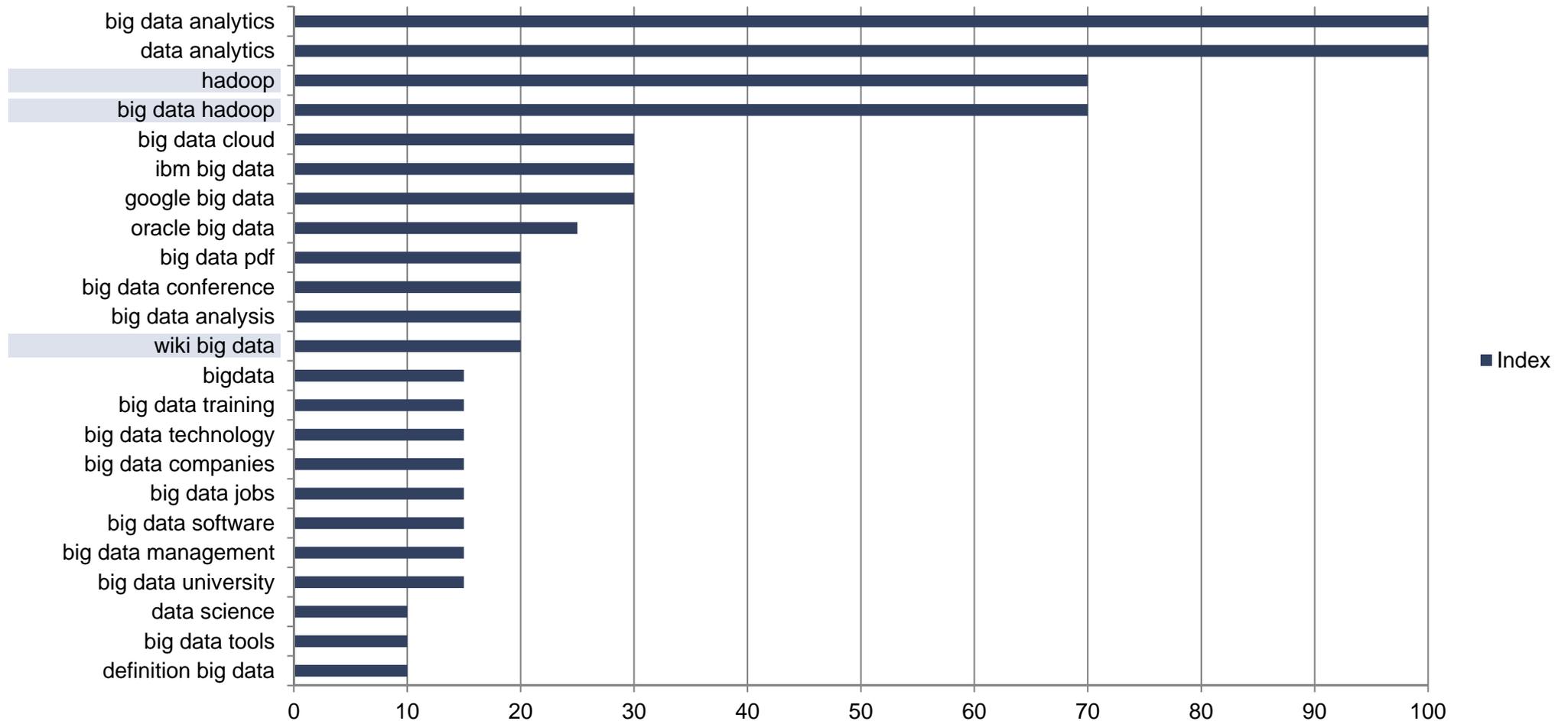


Plateau will be reached in:

- less than 2 years
- 2 to 5 years
- 5 to 10 years
- ▲ more than 10 years
- ⊗ obsolete before plateau

2 Die häufigsten Suchkombinationen

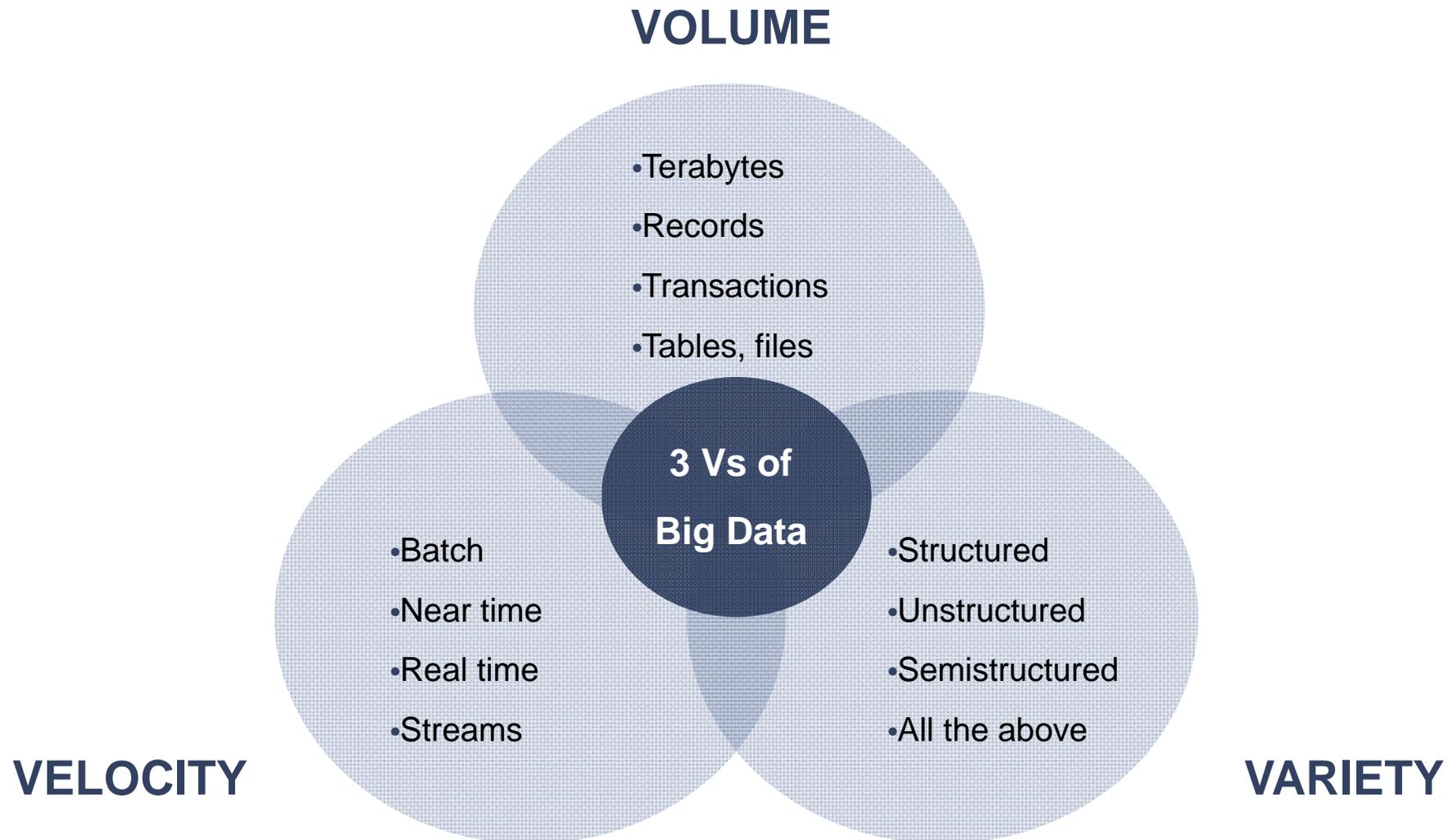
Die 25 meistgesuchten Begriffe im Zusammenhang mit „Big Data“, sortiert nach ihrer relativen Häufigkeit



[Google Trends, eigene Recherche, 12.09.2013]

2 Big Data Analytics (Definition)

The three Vs of big data



[TDWI 2011]

2 Big Data Analytics (Definition)

„Big Data bietet Methoden und Technologien für das Erfassen, Speichern und Analysieren **poly-strukturierter** Daten genau dort, wo klassische analytische Informationssysteme heute an ihre Grenzen stoßen.“

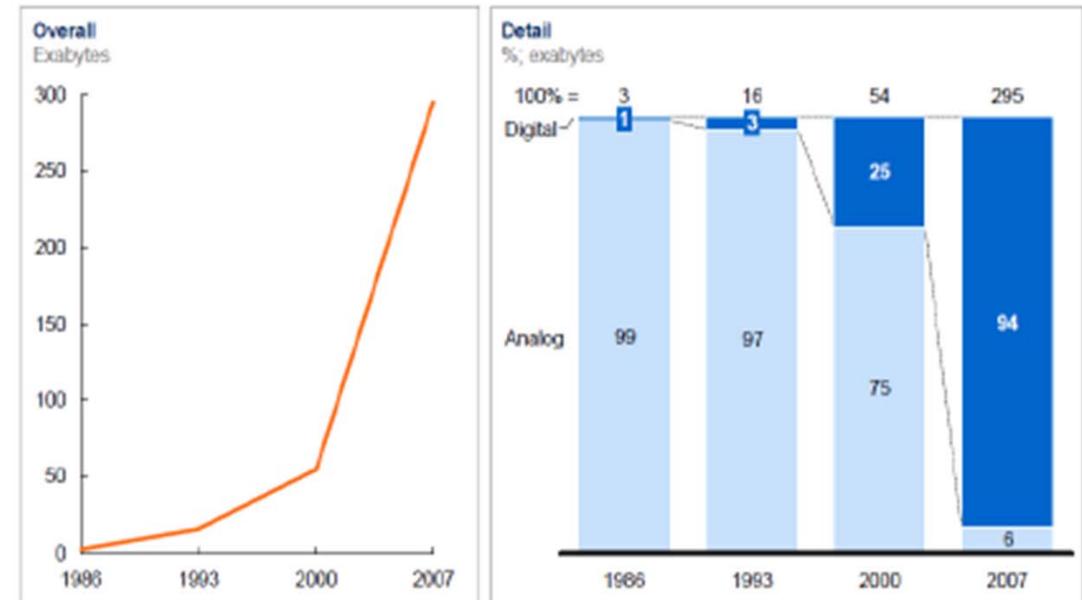
[BARC in Computerwoche (09/12)]

2 Big Data und Volumen

- As introduced, the amount of data to be analyzed is constantly growing
- In this situation a sole concentration on manual / interactive analysis methods like table or OLAP reports is not sufficient anymore
- Methods and tools that (semi-) automatically generate knowledge from large data sets and documents are needed

Data storage has grown significantly, shifting markedly from analog to digital after 2000

Global installed, optimally compressed, storage

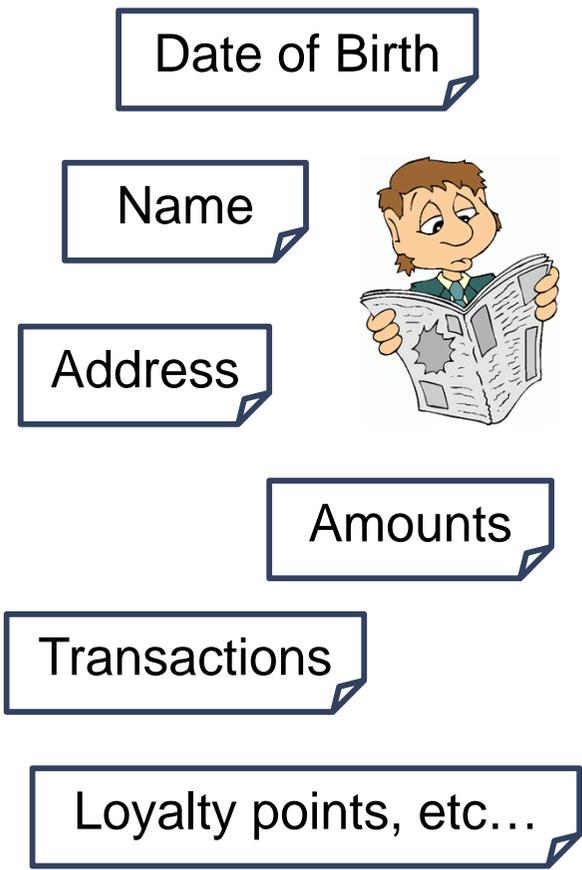


NOTE: Numbers may not sum due to rounding

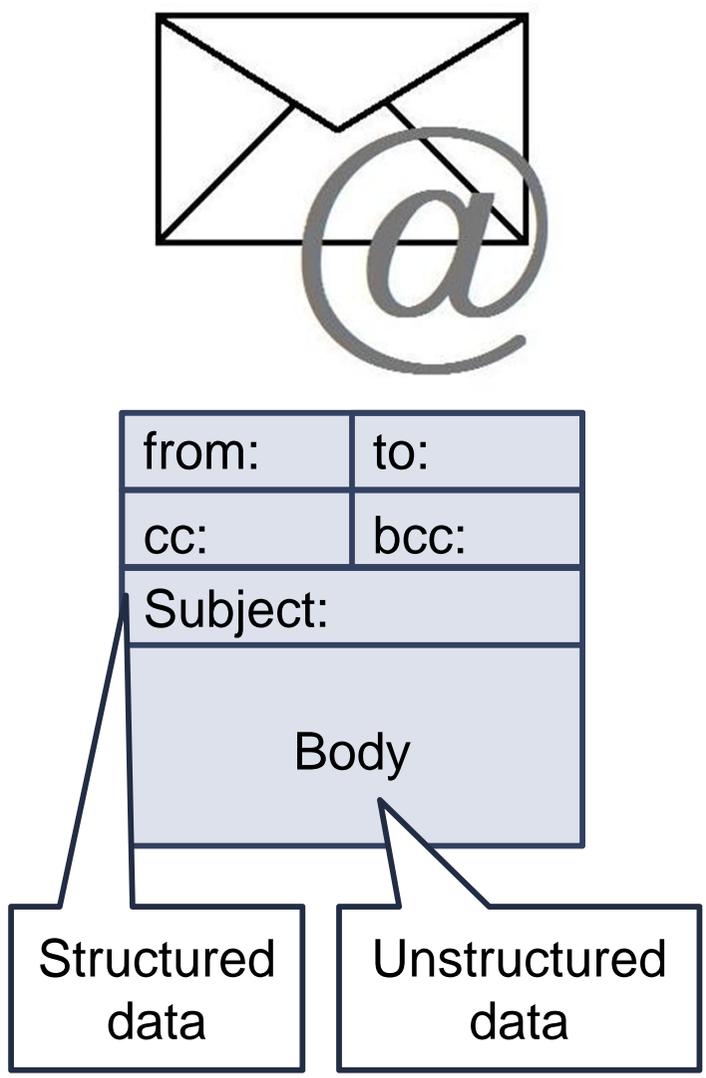
SOURCE: Hilbert and López, "The world's technological capacity to store, communicate, and compute information," *Science*, 2011

2 Big Data und Vielfalt

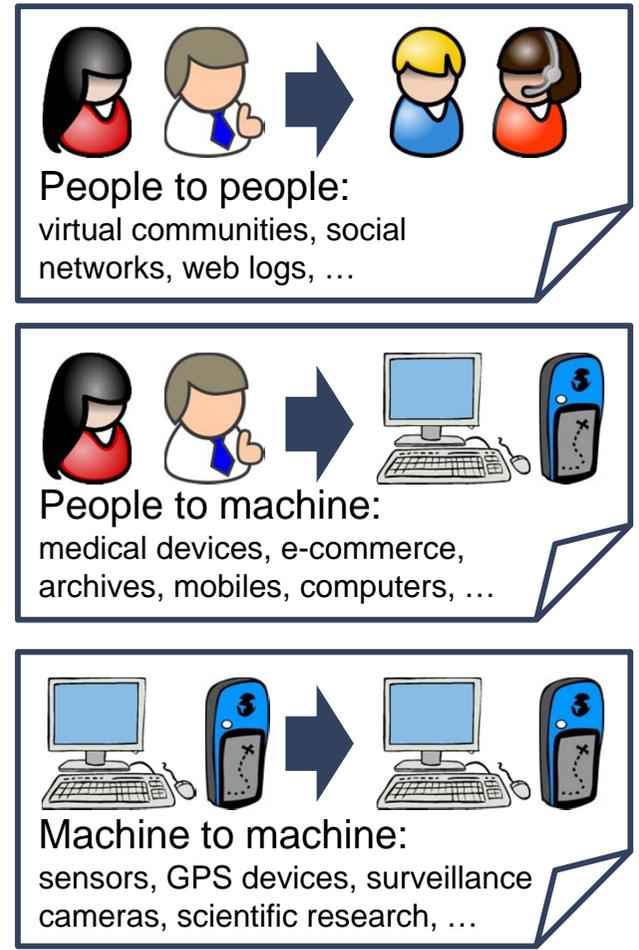
Structured data



Semi-structured data



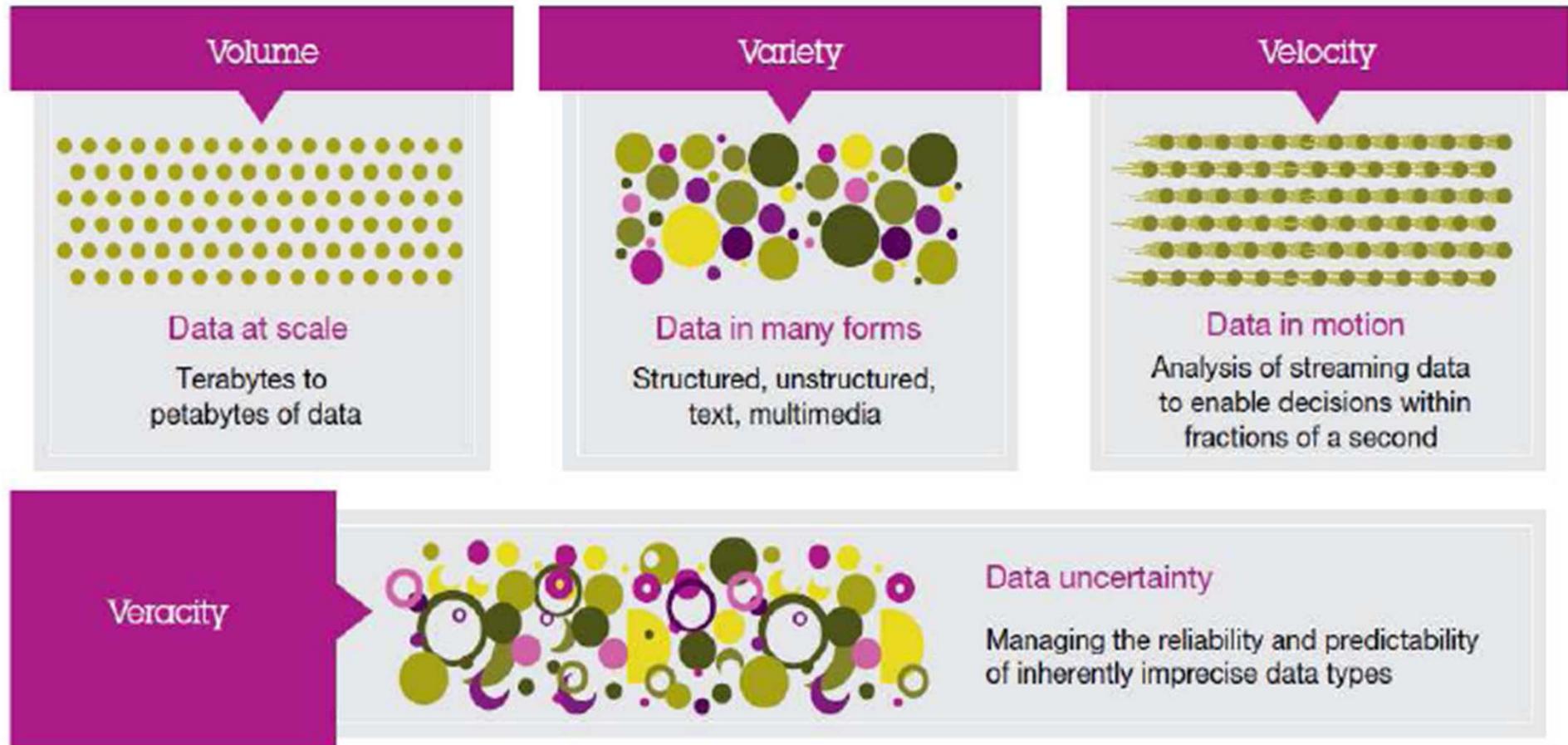
Unstructured data



[Klein et al. 2013]

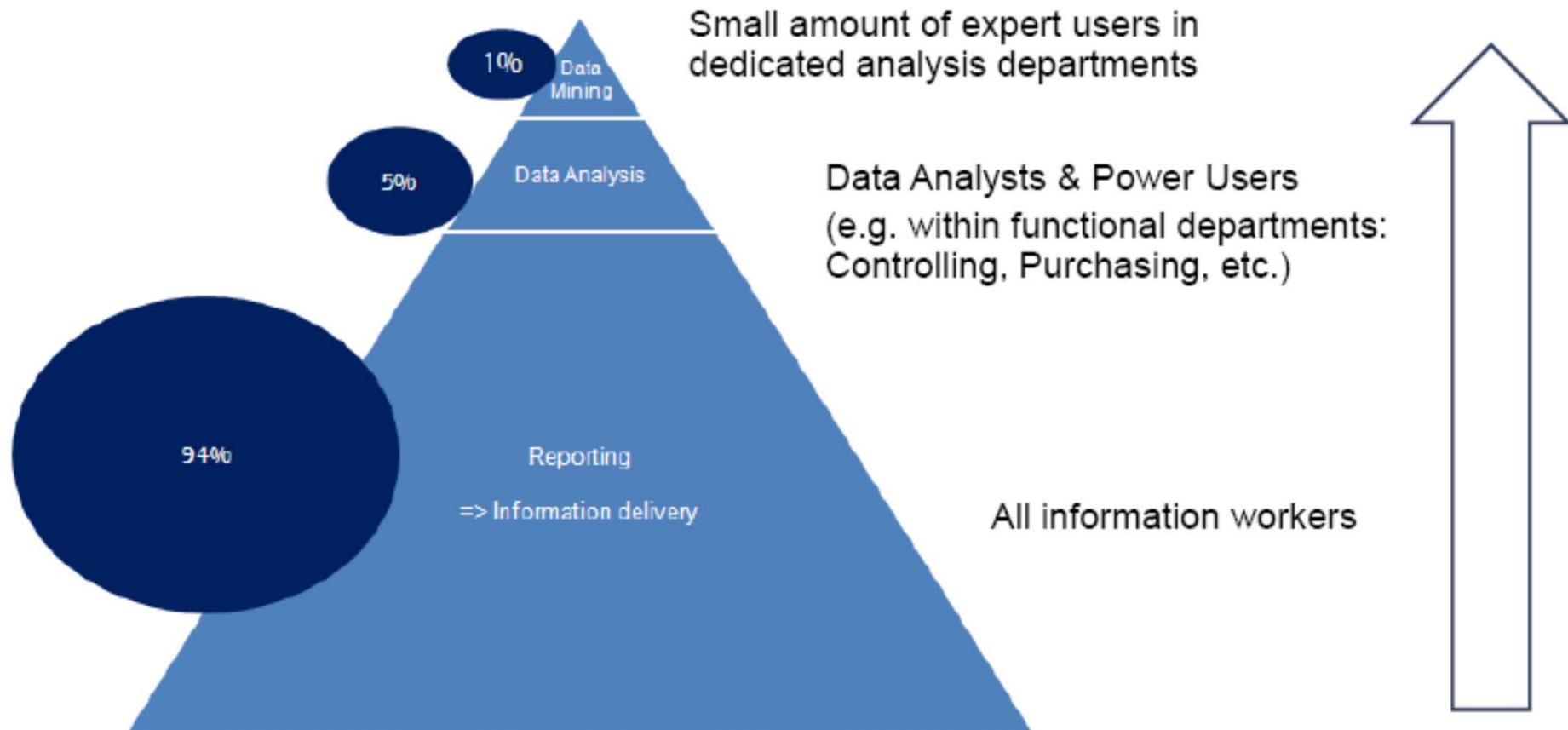
Grafiken: [<http://www.martinprint.com.au>, www.clker.com]

2 Big Data und Vertrauen



Schroeck et al. (2012) – IBM Institute for Business Value

2 Big Data Nutzer



*proportions as of 2011

Gartner (2007-2012) *Gartner EXP Worldwide Survey of More than 1,500 CIOs

2 Big Data Enabling Technologies

Technologies

- Volume → Clustering
- Velocity → Data Processing
- Variety → New forms of (streaming) Analytics

Techniques (Veracity: very important discussion)

- Cognitive Decision making and Decision Support
- Skills: Human Resources and Education (“Data Scientist”)
- Organization change: The Data-Driven Company

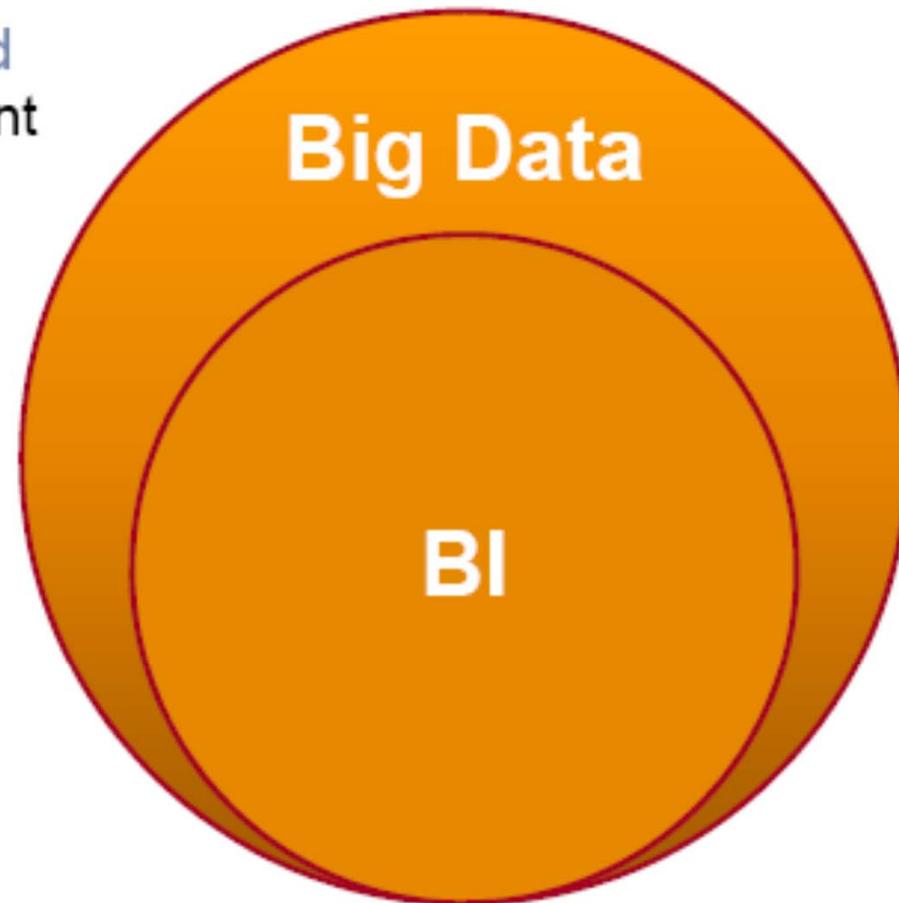
2 Schwerpunkte analytischer Systeme und Big Data

Analytische Systeme	Big Data
Zentrale Datenhaltung, alle Daten müssen exakt zueinander passen	Daten existieren an mehreren Stellen, Ungenauigkeiten sind akzeptabel
Qualitativ hochwertige Daten	Einfachheit der Nutzung
Strukturierte, bereinigte und aggregierte Daten	Verarbeitung der Rohdaten mit vielen unterschiedlichen Formaten
Wiederkehrende Berichte	Interaktion in Echtzeit
Periodische Erstellung	Optimiert für Flexibilität
Zentralistische Organisation	Heterogene, dezentrale Organisation

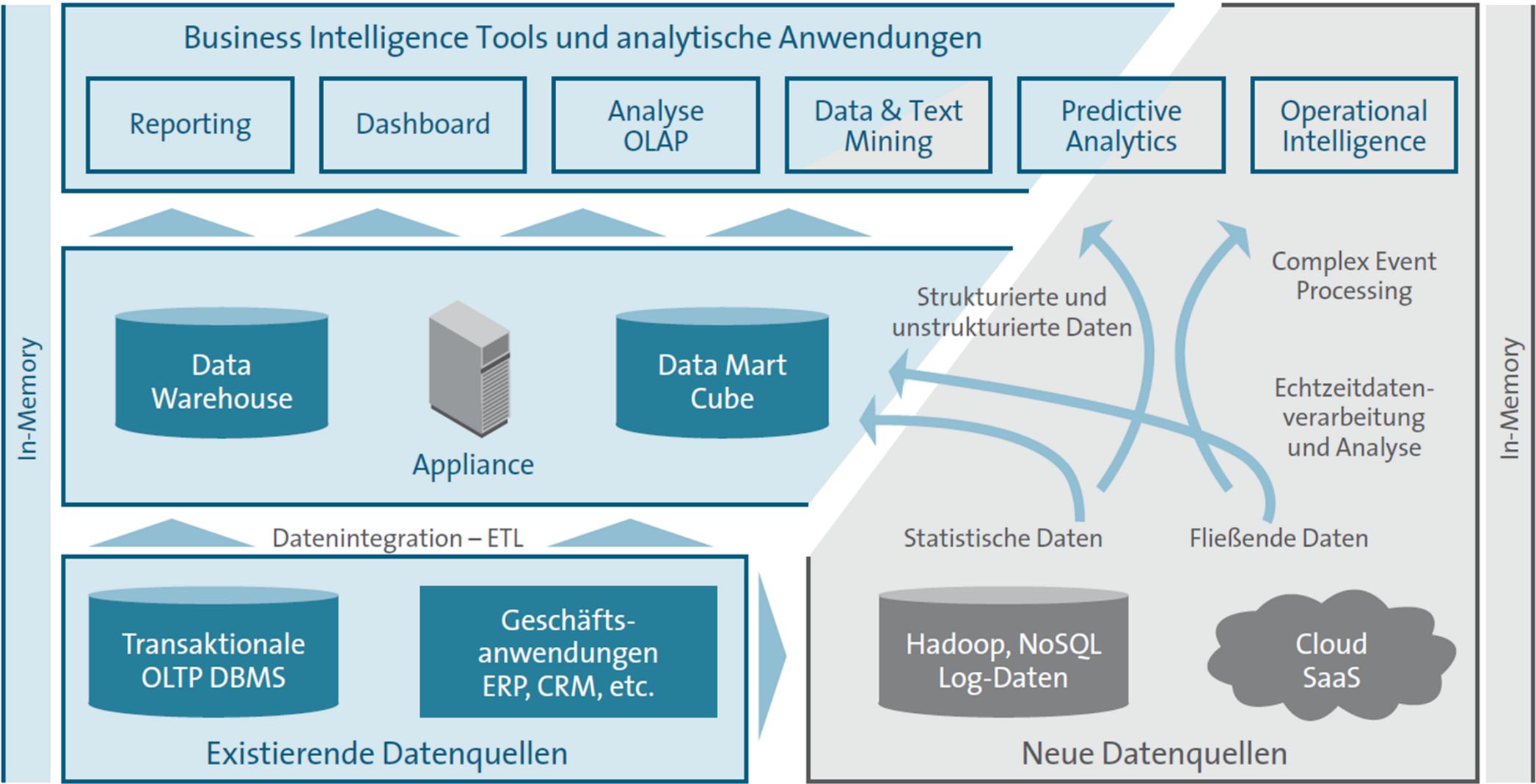
[BITKOM 2013]

2 BI und Big Data

- Incorporates **new technologies and techniques** – that extend the current scope of BI
- Makes **data accessible** that is unstructured and of heterogeneous sources
- Big Data Analytics (**Data Scientist and Data Journalist**)
- Deeply grounded in Corporate **Decision Making**
- Organizational challenge: “**Data-Driven Company**”



2 Integrierte Anwendungslandschaft



[BITKOM 2013]

2 Types of analytical platforms (1/2)

Technology	Description	Vendor/Product
Massively parallel processing analytics databases	Row-based databases designed to scale out on a cluster of commodity servers and run complex queries in parallel against large volumes of data	Teradata Active Data Warehouse, Greenplum (EMC), Microsoft Parallel Data Warehouse, Aster Data (Teradata), Kognitio, Dataupia
Columnar databases	Database management systems that store data in columns, not rows, and support high data compression ratios	ParAccel, Infobright, Sand technology, Sybase IQ (SAP), Vertica (Hewlett-Packard), 1010data, Exasol, Calpont
Analytical appliances	Preconfigured hardware-software systems designed for query processing and analytics that require little tuning	Netezza (IBM), Teradata appliances, Oracle Exadata, Greenplum Data Computing Appliance (EMC)
Analytical bundles	Predefined hardware and software configurations that are certified to meet specific performance criteria, but customers must purchase and configure themselves	IBM SmartAnalytics, Microsoft FastTrack

[Eckerson 2011]

2 Types of analytical platforms (1/2)

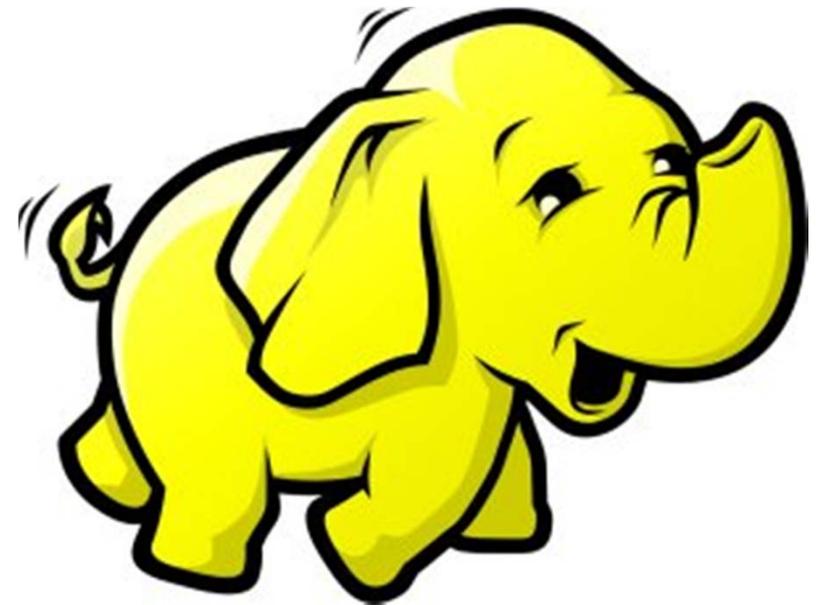
Technology	Description	Vendor/Product
In-memory databases	Systems that load data into memory to execute complex queries	SAP HANA, Cognos TM1 (IMB), QlikView, Membase
Distributed file-based systems	Distributed file systems designed for storing, indexing, manipulating and querying large volumes of unstructured and semi-structured data	Hadoop (Apache, Cloudera, MapR, IBM, HortonWorks), Apache Hive, Apache Pig
Analytical services	Analytical platforms delivered as hosted or public-cloud-based services	1010data, Kognition
Nonrelational	Nonrelational databases optimized for querying unstructured data as well as structured data	MarkLogic Server, MongoDB, Splunk, Attivio, Endeca, Apache Cassandra, Apache Hbase

[Eckerson 2011]

2 Hadoop

Was ist Hadoop?

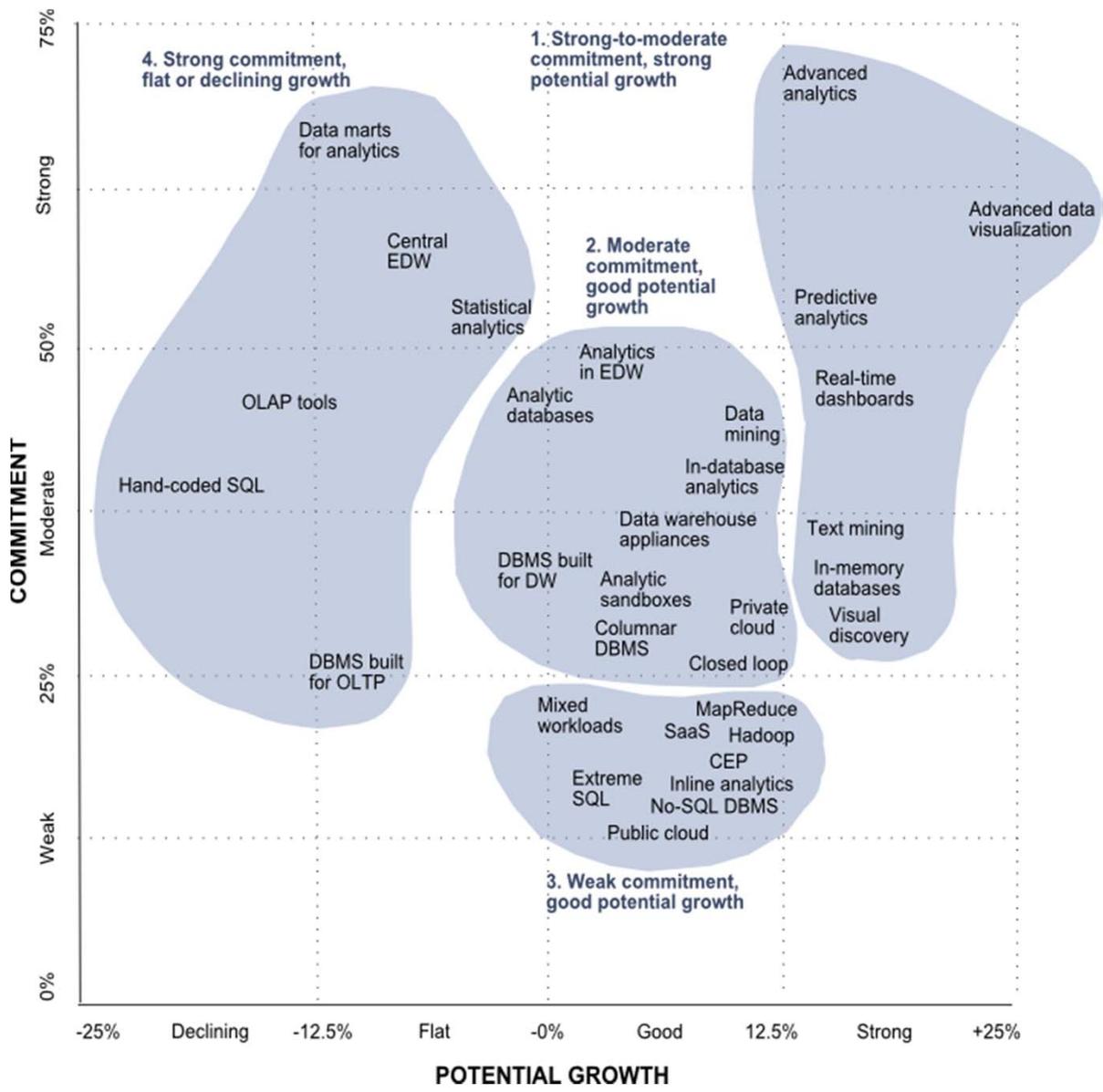
- Freies, Java-basiertes Framework
- Skalierbare, verteilt arbeitende Software
- Basiert auf MapReduce-Algorithmus von Google und Google-Dateisystem
- Ermöglicht Rechenprozesse auf verteilten Computersystemen
- Existiert in verschiedenen Distributionen von verschiedenen Herstellern (ähnlich Unix/Linux)
- Hadoop wurde nach dem Spielzeugelefanten des Sohns des Entwicklers benannt. Der Elefant ist heute das Symbol von Hadoop.



Agenda

- 1 Einleitung
- 2 Big Data
- 3 Empirische Studien
- 4 Markt, Tools und Anwendungen
- 5 Zusammenfassung und Ausblick

3 Unternehmensinschätzungen (nach TDWI)



Options for Big Data Analytics

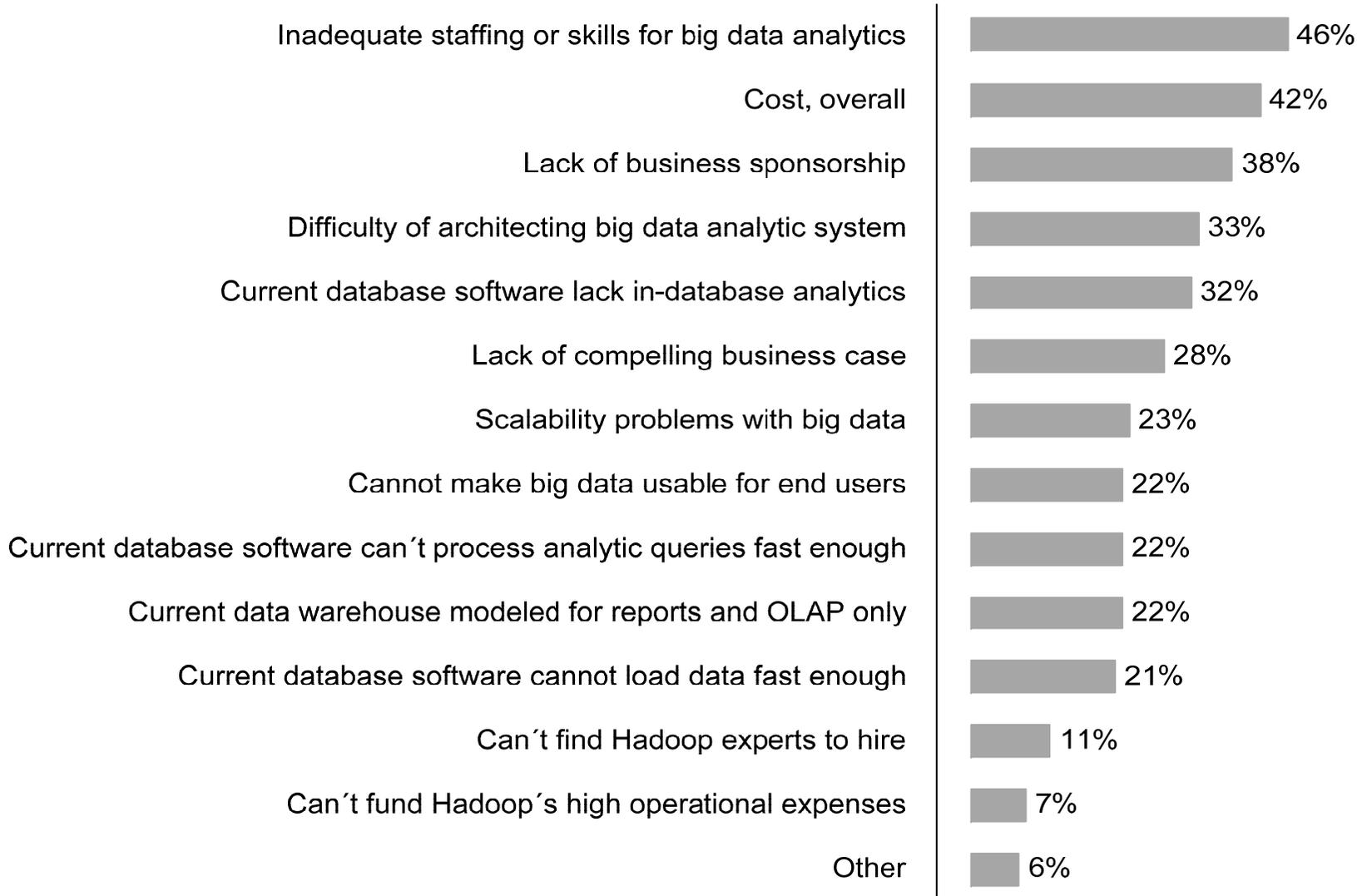
Plotted by Potential Growth and Commitment

[TDWI 2011]

3 Unternehmenseinschätzungen (nach TDWI)



In your organization, what are the top potential barriers to implementing big data analytics? (Select five or fewer.)

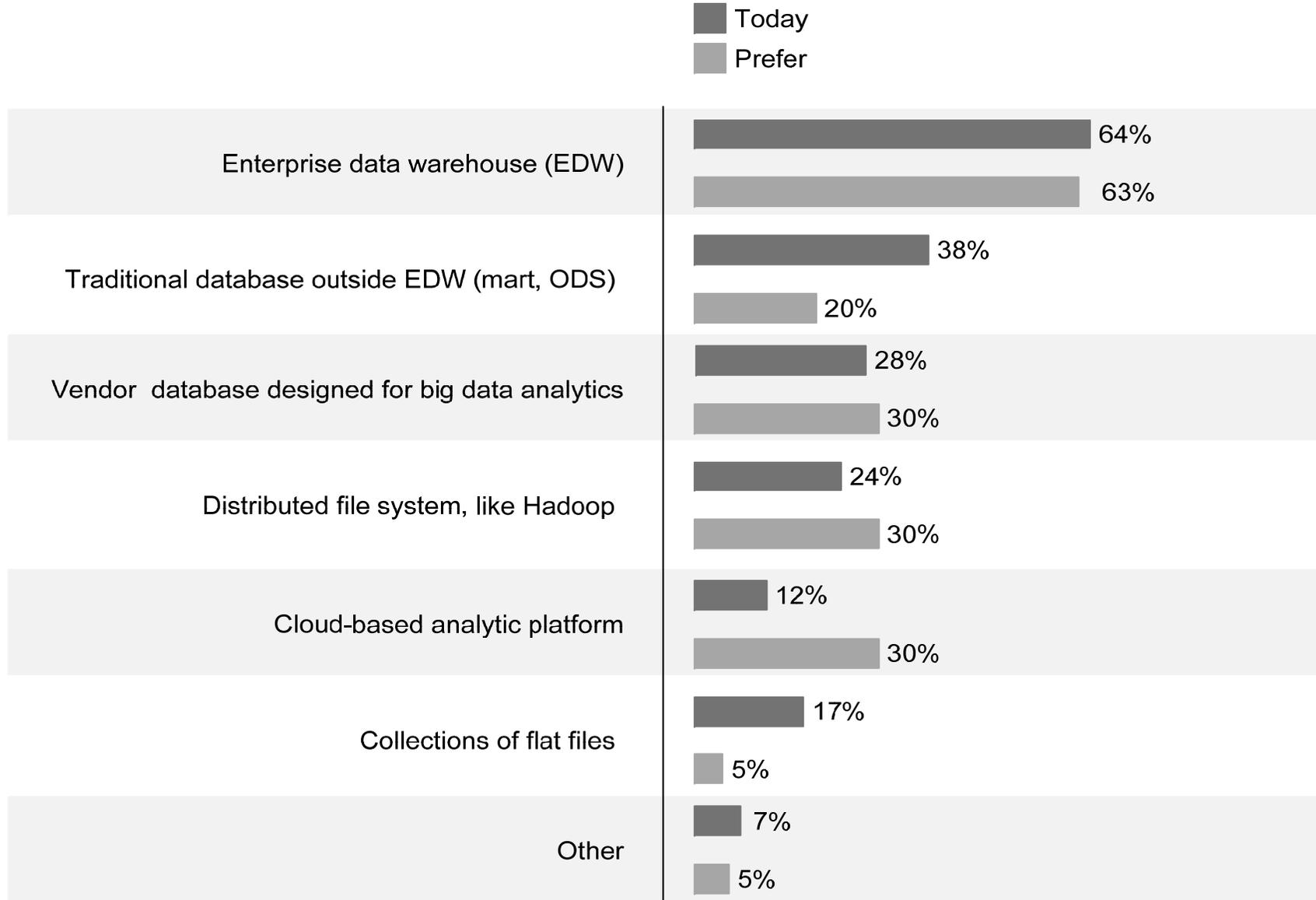


[Russom 2011]

3 Unternehmenseinschätzungen (nach TDWI)



**Today, where is big data for advanced analytics managed and operated on?
Where would you prefer that big data for advanced analytics be managed and operated on?**

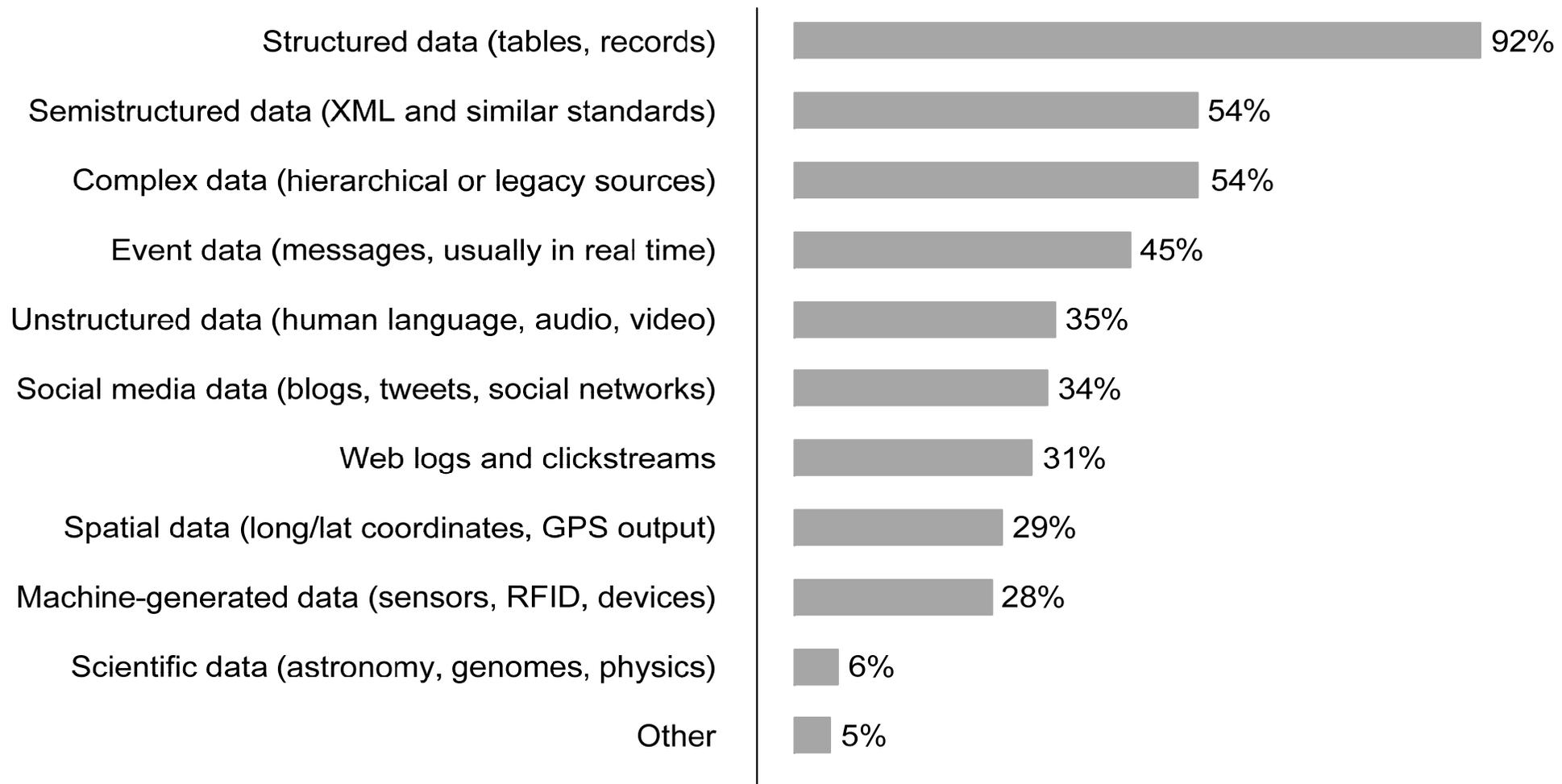


[Russom 2011]

3 Unternehmenseinschätzungen (nach TDWI)



Which of the following data types are you collecting as big data and/or using with advanced analytics today? Select all that apply.

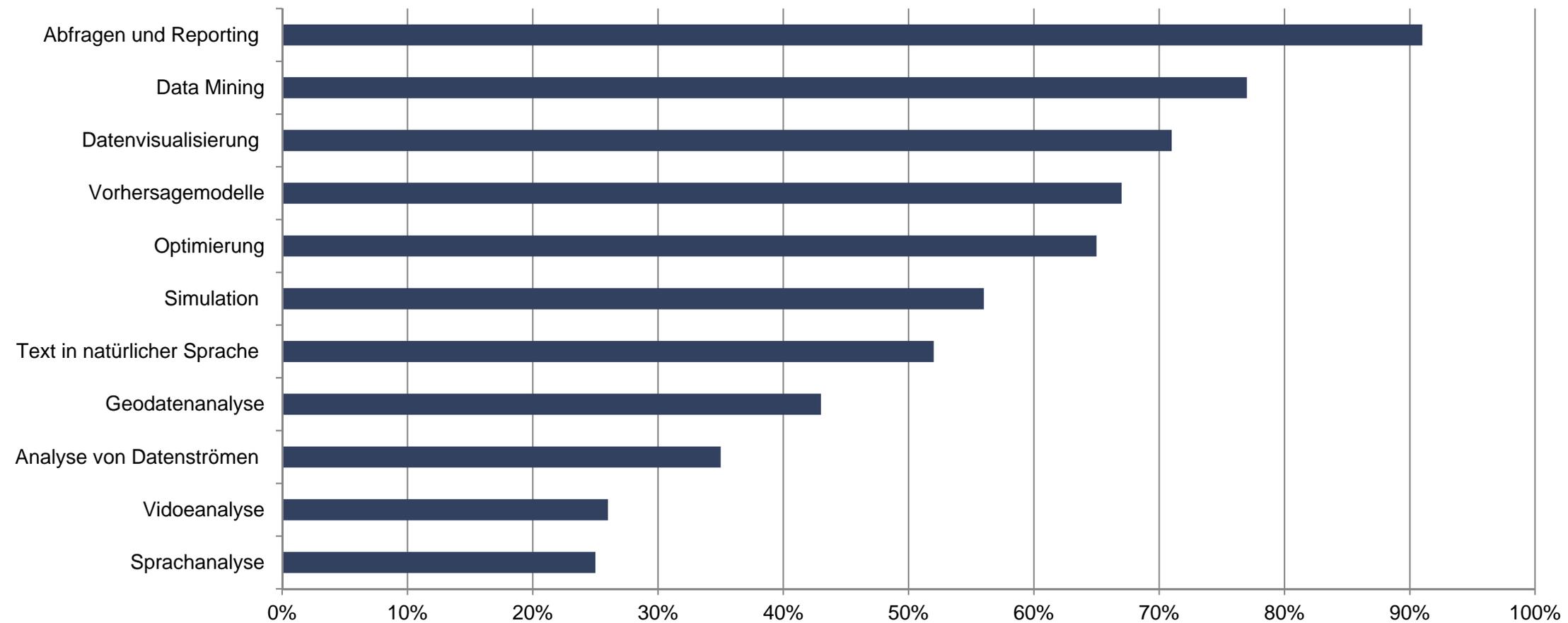


[Russom 2011]

3 Unternehmenseinschätzungen (nach IBM)



Welche Analysetools stehen Ihnen für eine Big-Data-Initiative zur Verfügung?

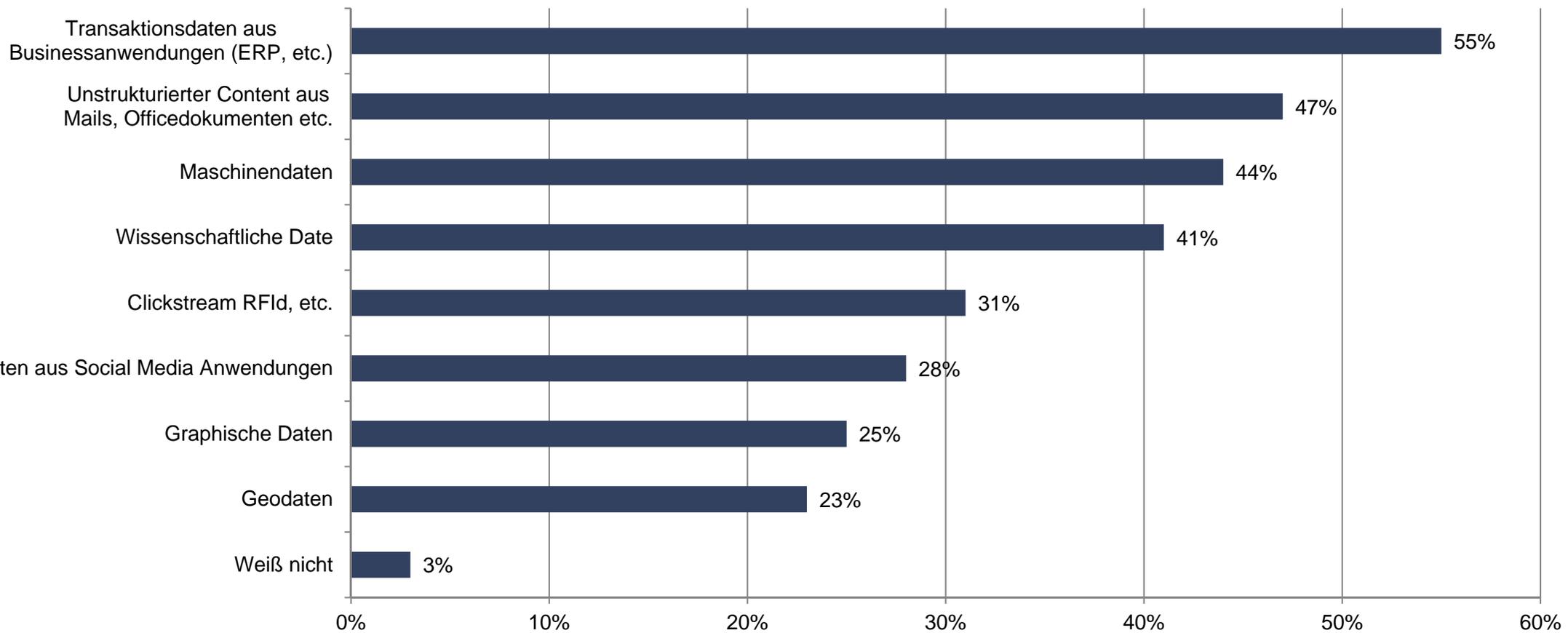


[IBM 2012]

3 Unternehmenseinschätzungen (nach IDC)



Aus welchen Datenquellen nutzen Sie Daten für Big Data Analytics?

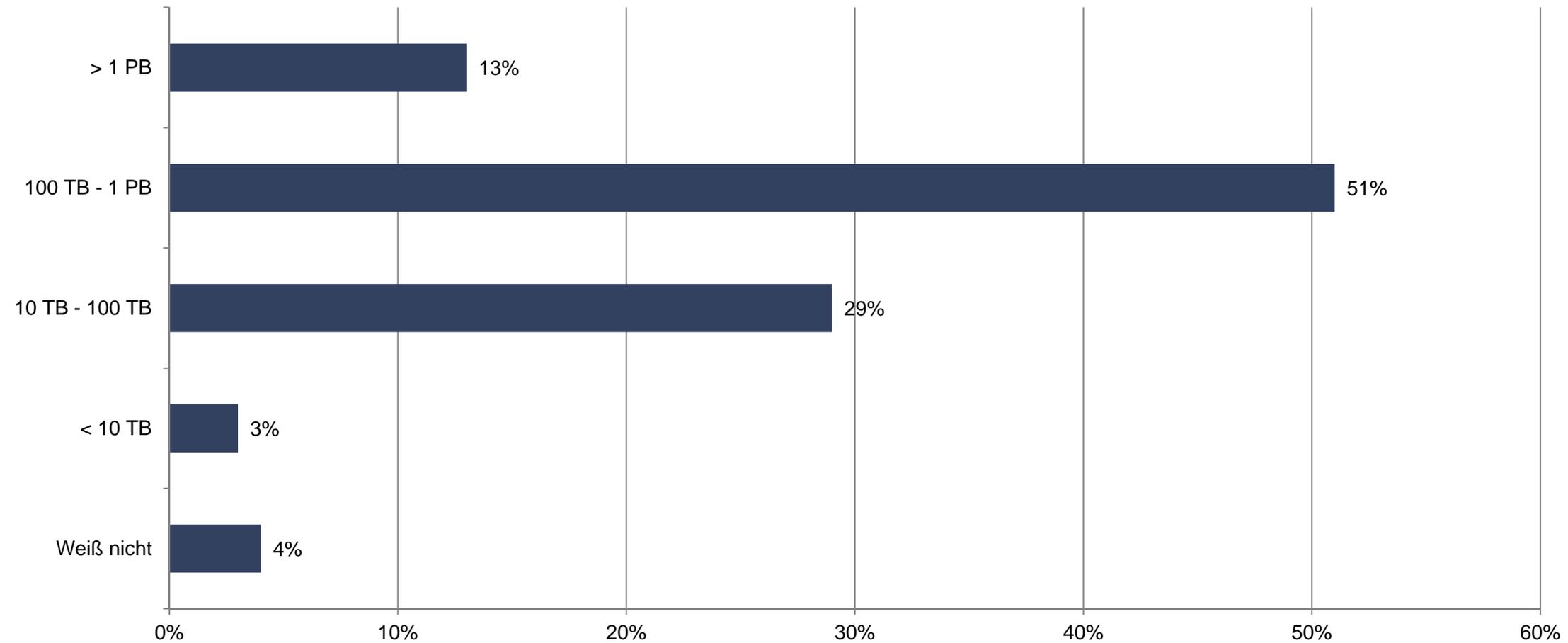


[n=150, IDC 2012]

3 Unternehmenseinschätzungen (nach IDC)



Welches Datenvolumen im Businesscase betrachten Sie als Big Data?

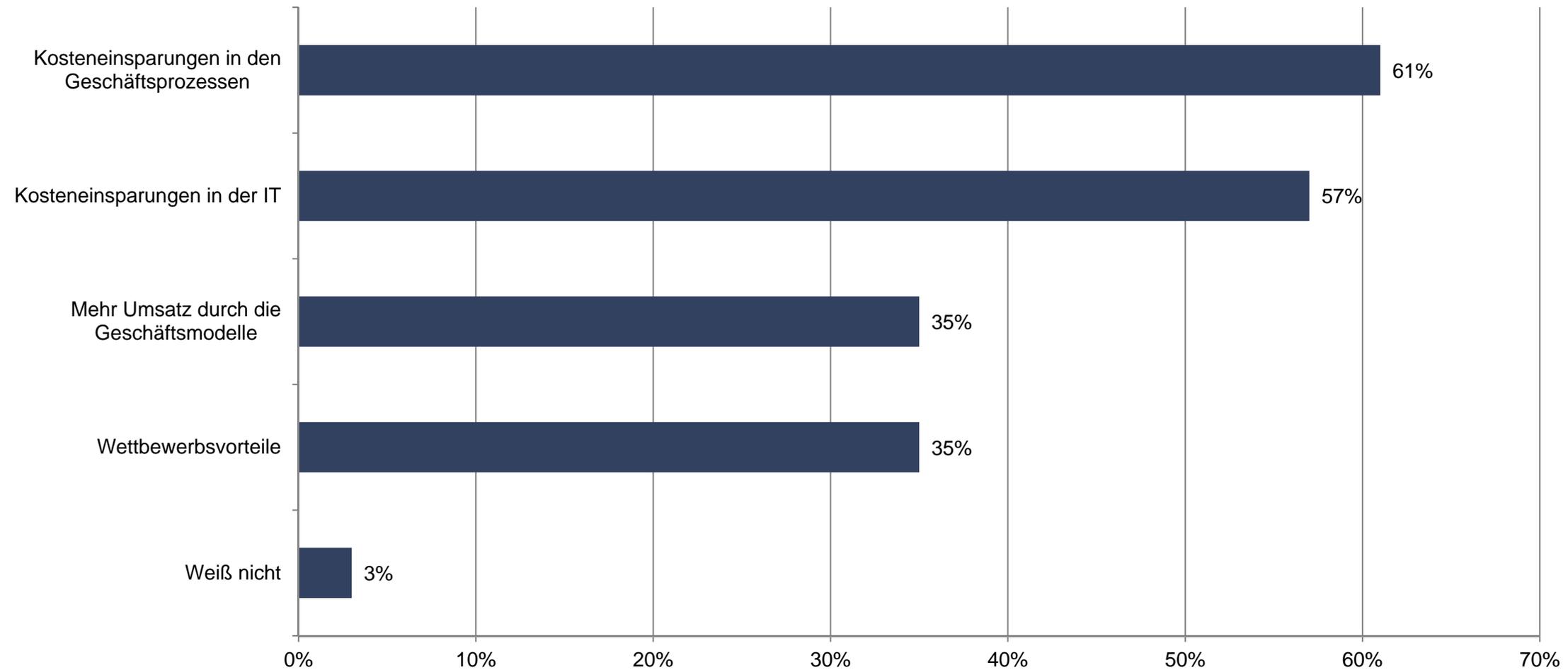


[n=150, IDC 2012]

3 Unternehmenseinschätzungen (nach IDC)



Welchen Nutzen erwarten Sie durch den Einsatz von Big Data Analytics?

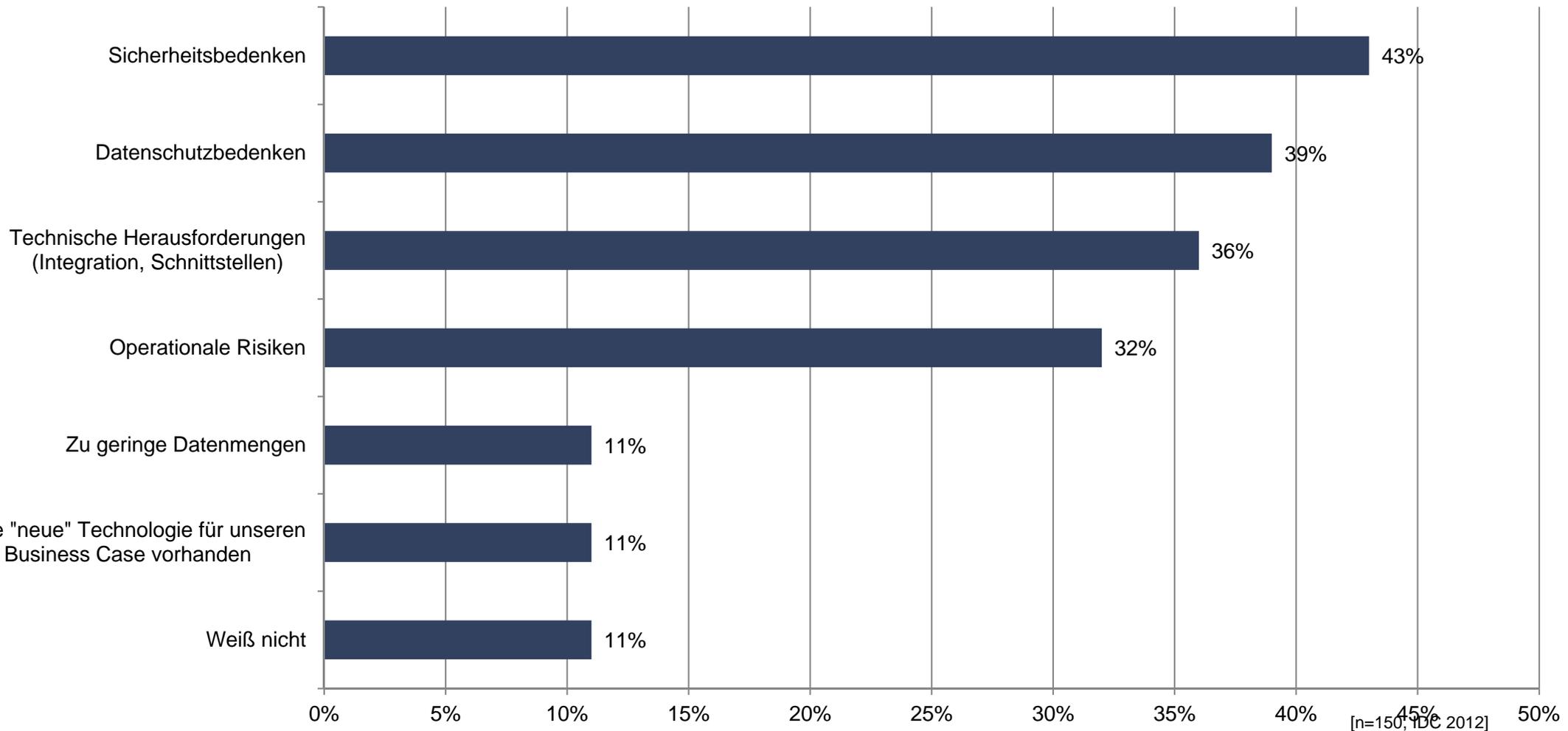


[n=150, IDC 2012]

3 Unternehmenseinschätzungen (nach IDC)

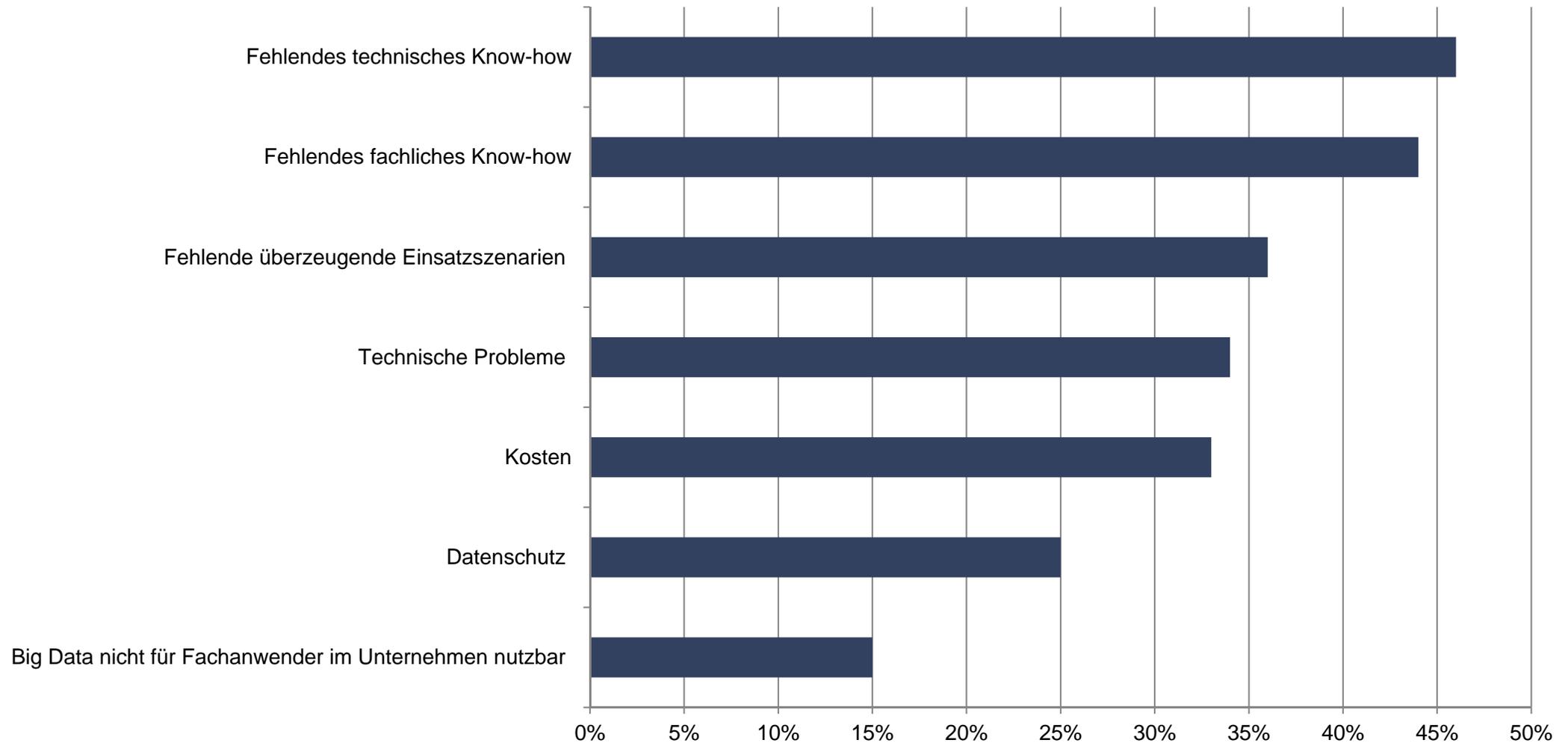


Welche Argumente sprechen gegen den Einsatz von Big Data Analytics Lösungen?



3 Unternehmenseinschätzungen (nach BARC)

Welche Probleme sehen Sie beim Einsatz von Big Data ?

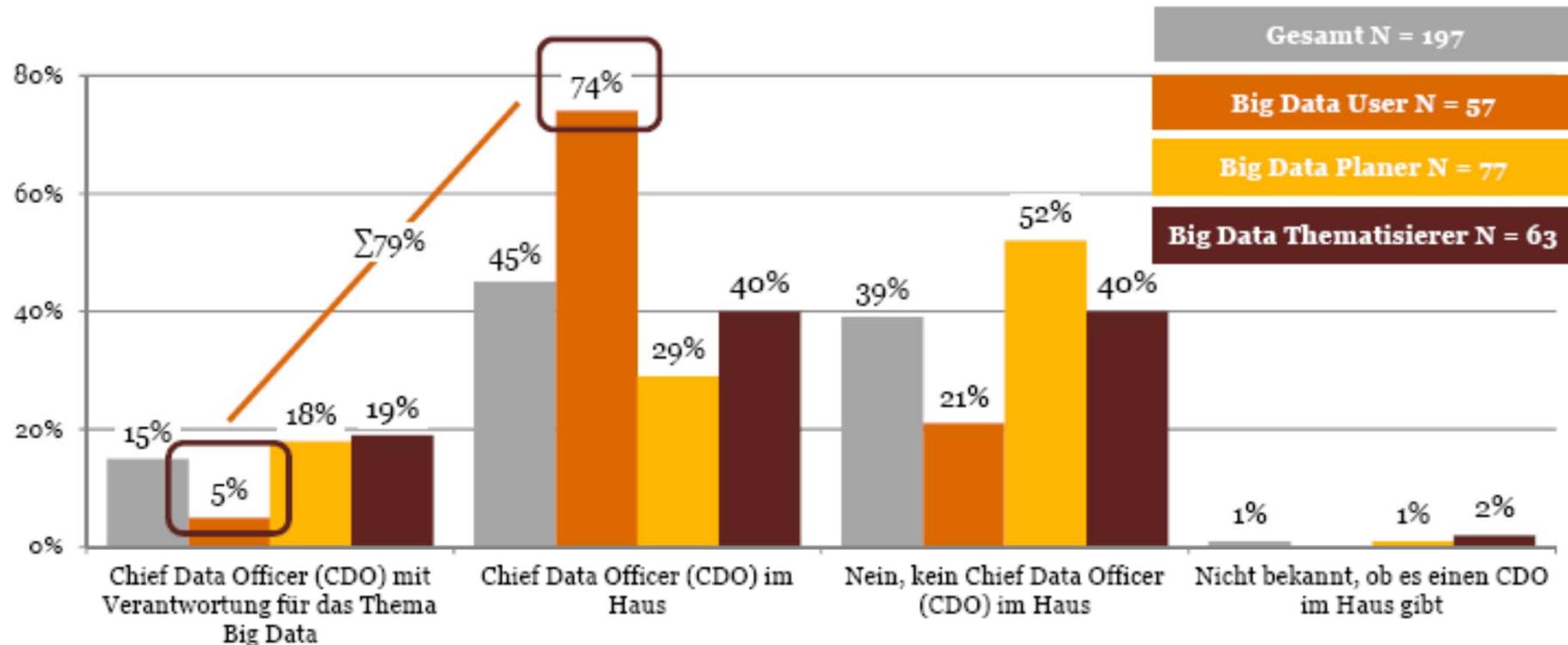


[BARC, 2013]

3 Unternehmenseinschätzungen (nach PwC)

Ergebnisse der PwC Big Data Studie 2014

CDOs haben nur selten die Verantwortung für Big Data

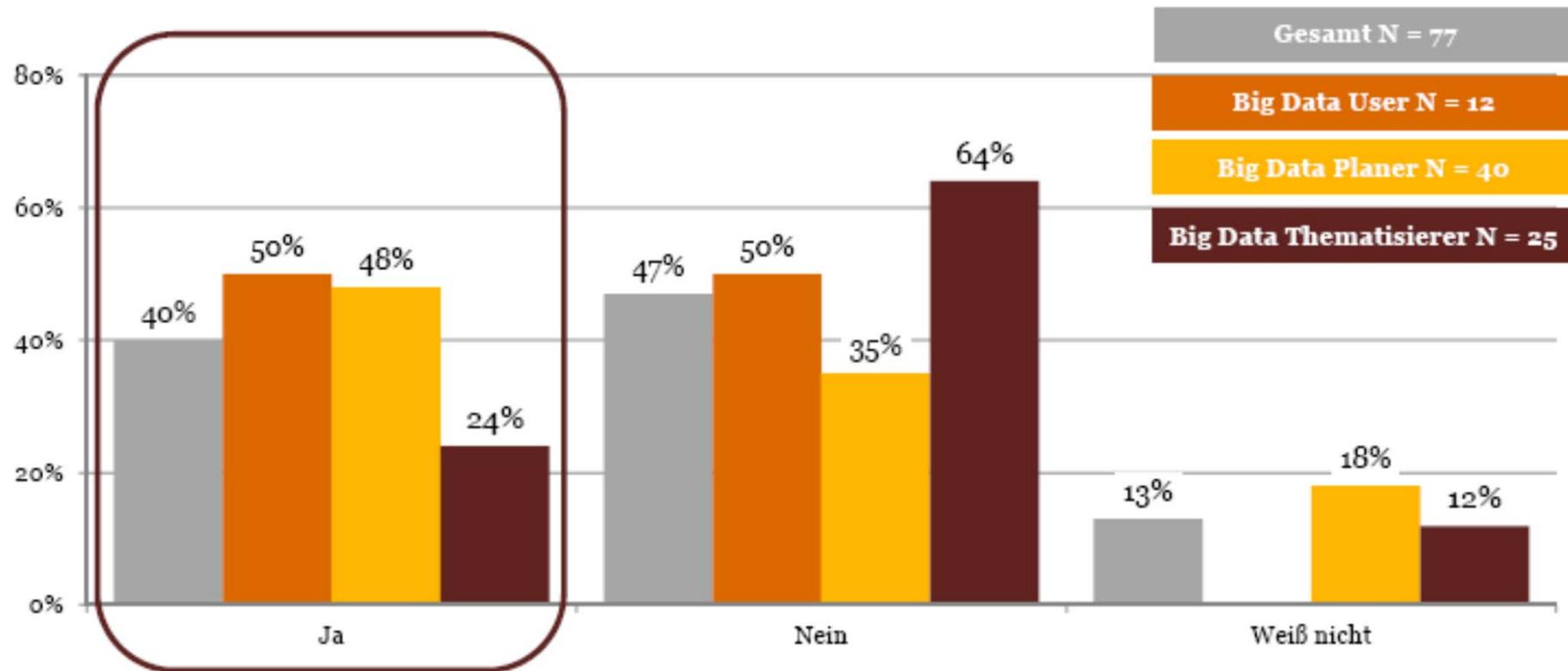


F4/5a: Wer ist in Ihrem Unternehmen für das Thema Big Data hauptsächlich verantwortlich? / Gibt es in Ihrem Unternehmen die Stelle „Chief Data Officer“ (CDO)?
(Einfachnennung, an Big Data User, Planer und Thematisierer, N = 197)

3 Unternehmenseinschätzungen (nach PwC)

Ergebnisse der PwC Big Data Studie 2014

40 Prozent der Industrieunternehmen, die noch keinen Chief Data Officer haben, planen die Einrichtung einer solchen Stelle.



F5: Plant Ihr Unternehmen derzeit die Einrichtung einer Chief Data Officer-Stelle (CDO)? (Einfachnennung, an Big Data User, Planer und Thematisierer, N = 77, die noch keine CDO-Stelle im Unternehmen haben)

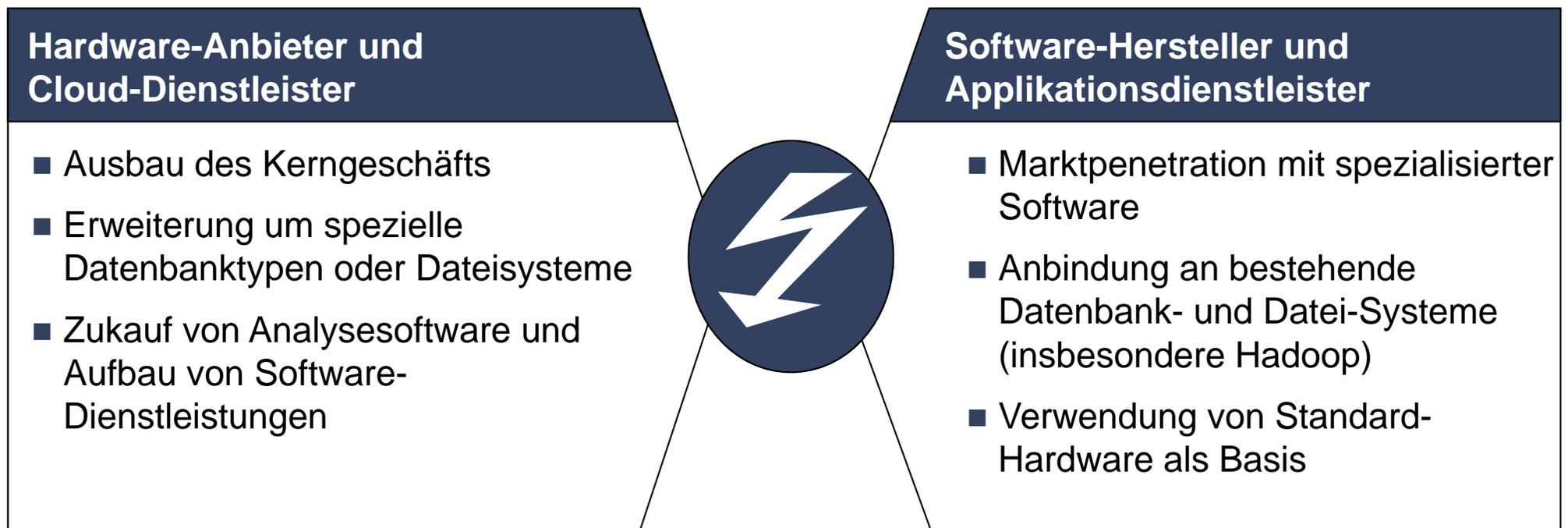
Agenda

- 1 Einleitung
- 2 Big Data
- 3 Empirische Studien
- 4 Markt, Tools und Anwendungen
- 5 Zusammenfassung und Ausblick

4 Anbieter-Entwicklung bei Big Data

Je nach Herkunft bewegen sich die Anbieter von unterschiedlichen Richtungen hin zu einer integrierten Lösung

Bis 2012 wurden ca. 15 Milliarden Dollar für den Kauf von spezialisierten Informationsdienstleistern aufgewendet



[vgl. Storage Consortium]

4 Anbieter-Entwicklung bei Big Data

Der Ausbau der Big-Data-Sparte findet sich bei allen großen Anbietern, beispielhaft genannt seien hier:

IBM

- InfoSphere BigInsights, Netezza Data Warehouse Appliances, u.a.: Integrierte Lösungen, verschiedene Wahlmöglichkeiten für Hard- und Software, diverse Analysesoftware

Oracle

- Big Data Appliance: Integrierte Lösung aus Hard- und Software (Oracle Linux/NoSQL, Apache Hadoop, Cloudera Manager)

EMC

- Atmos, Isilon, Greenplum, Pivotal HD u.a.: Hardware- und Software-Komponenten, spezialisierte Datenbanken und Analysesoftware

Eigene Recherche nach Anbieter-Informationen

4 Anwendungen

Sensors deployed on buoys in Galway Bay

- Stream data on ocean conditions and water quality in real-time
- Data from buoys **combined** with other online databases, e.g. geographical information
- Creating a **“smart ocean” cluster**:
 - Alternative energy developers access real-time wave data to determine effectiveness of prototype **wave-energy generators**
 - Fishermen can use environmental data to know when to put to sea
 - Fishery managers can monitor and track **water quality issues**
 - Help increase **public safety**



4 Anwendungen

- Coca-Cola uses Big Data to produce orange juice that has a **consistent taste** year-round, although the oranges used have a peak-growing season of just **three months**.
- They developed an **algorithm**, called the **Black Book model**, that combines various **data sets** to create a consistent taste such as
 - satellite imagery
 - weather data
 - expected crop yields
 - regional consumer preferences
 - detailed data about the myriad of 600 different flavors that make up an orange



4 Anwendungen

Anwender	Anbieter	Problemstellung	Volumen
 arvato Systems / Deutschland Card	arvato Systems GmbH	<ul style="list-style-type: none"> • Neue Teilnehmer im DeutschlandCard-Programm führen zu mehr als 4 GB/s Datendurchsatz im DWH. • Die bisherige Landschaft soll nur ergänzt werden. 	<ul style="list-style-type: none"> • Datenwachstum im dreistelligen GB-Bereich pro Monat • Auswertungen immer tagesaktuell und maximal im Minutenbereich
 dm, Geschäftsführung IT	Blue Yonder GmbH & Co. KG	<ul style="list-style-type: none"> • Mitarbeiterkapazitäten konnten in Sondersituationen nicht immer korrekt errechnet werden 	<ul style="list-style-type: none"> • Ermittlung aller Umsätze auf Tagesebene • Bis zu 450.000 Prognosen je Tag für alle Filialen
 Macy's	SAS Institute	<ul style="list-style-type: none"> • Anpassung von 270 Millionen Preispunkten (Artikel, Filiale, Umfeldkonstellation) • Nur wöchentliche Rechnungen möglich 	<ul style="list-style-type: none"> • Mehr als 2 TB je Analyse Rohdateneinsatz • Mehrfach untertägliche, filialindividuelle Preisermittlung
 Otto, Angebots- und Category Management Support	Blue Yonder GmbH & Co. KG	<ul style="list-style-type: none"> • Erhöhung der Lieferbereitschaft • Bestandsmanagements-optimierung • Prognoseerstellung 	<ul style="list-style-type: none"> • Täglich bis zu 135 GB • Jährlich über eine Milliarde Prognosen • Auswertung in Echtzeit

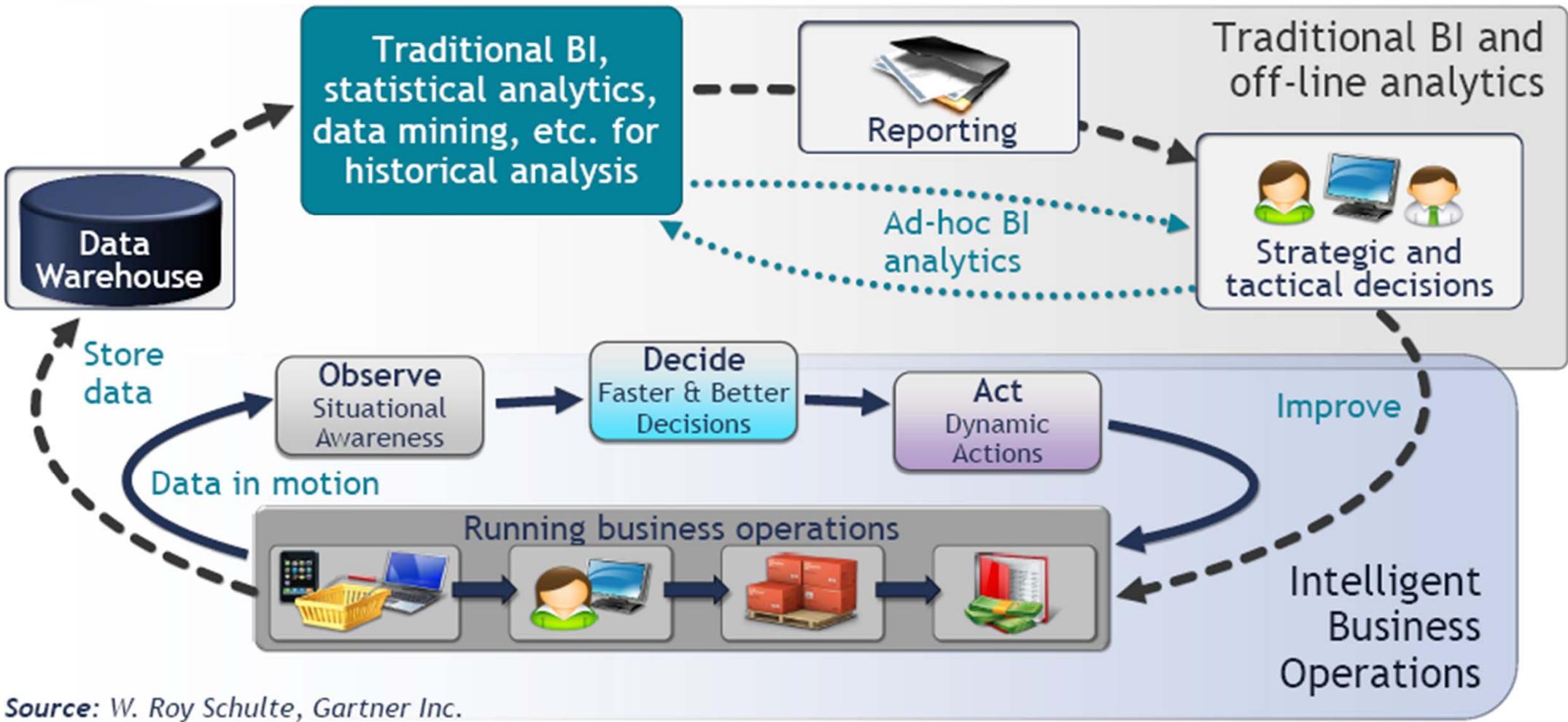
4 Anwendungen

Anwender	Anbieter	Problemstellung	Volumen
 Paymint AG	Fraunhofer IAIS	<ul style="list-style-type: none"> Weltweite Kosten bei Kreditkartenbetrug liegen bei mehr als 10 Mrd. Euro Fraud Management muss möglichst schnell werden 	<ul style="list-style-type: none"> Mehrere Milliarden Transaktionen pro Monat je Unternehmen Mustererkennung in Real-Time Schnelle Reaktion auf geänderte Betrugsstrategien
 Toll Collect GmbH	Toll Collect GmbH	<ul style="list-style-type: none"> Service-Level von 99,9% korrekter Mauttransaktionen Fehlerhafte Geräte müssen aus großen Datenmengen gefiltert werden 	<ul style="list-style-type: none"> 700.000 On-Board-Units im Einsatz 25 Mrd. gefahrene km/Jahr Mustererkennung in Echtzeit
 Vaillant, Group IT Consulting	SAP AG	<ul style="list-style-type: none"> Verwendung eines globalen, integrierten Systems Informationen auf Einzelproduktebene 	<ul style="list-style-type: none"> 1.200 Anwender 1 Mrd. Datensätze in DWH Geschwindigkeitserhöhung mit Faktor 4 bis 60
 XING AG	Exasol AG	<ul style="list-style-type: none"> Zeit für Datenverarbeitung eines Tages tlw. > 24 h Datenbanksystem heterogen und nicht skalierbar 	<ul style="list-style-type: none"> 12 Mio. Mitglieder 10 Mrd. Datensätze 30 TB Daten Echtzeit-Auswertungen möglich

Agenda

- 1 Einleitung
- 2 Big Data
- 3 Empirische Studien
- 4 Markt, Tools und Anwendungen
- 5 Zusammenfassung und Ausblick

5 Intelligent Business Operations



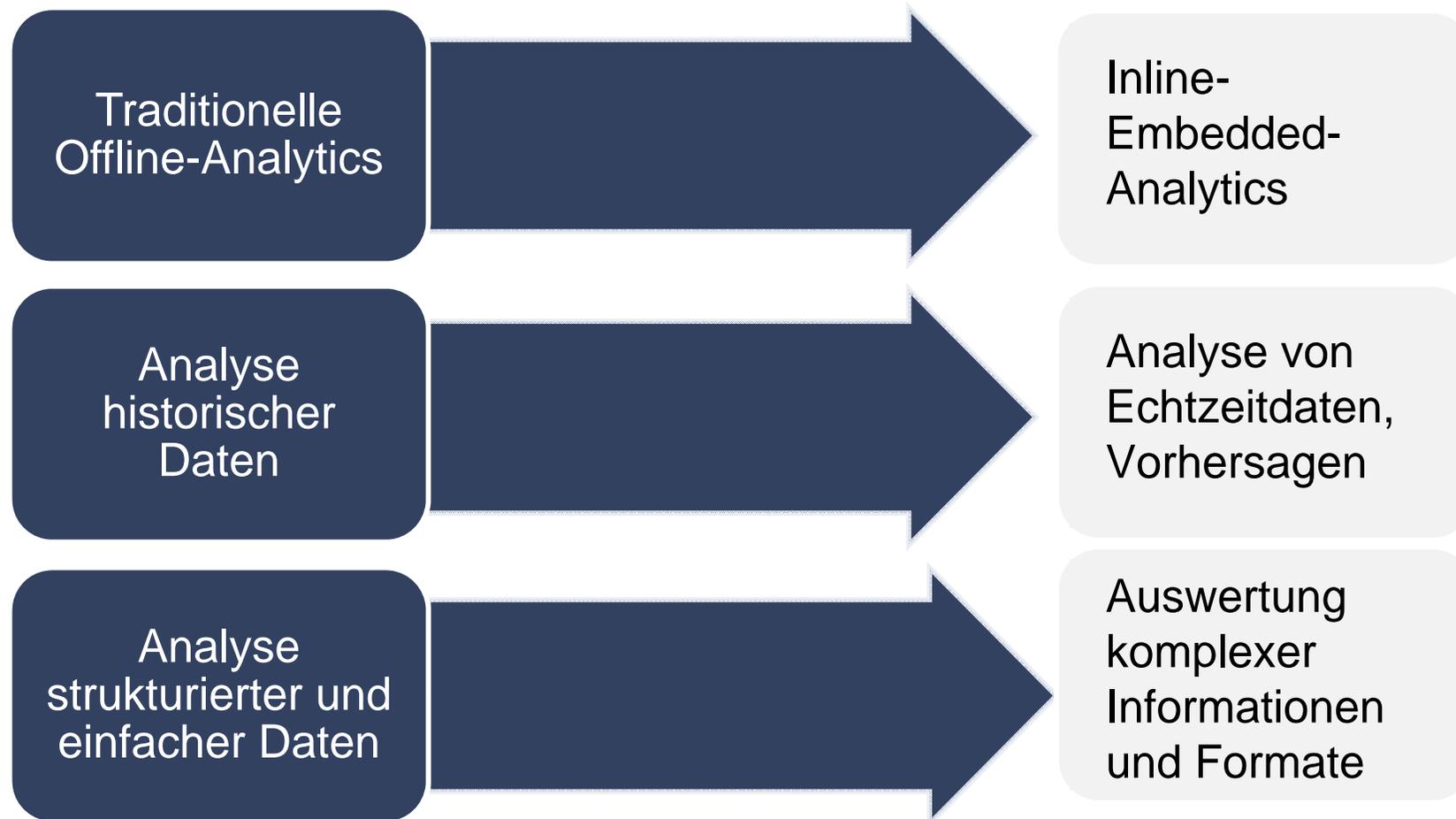
Source: W. Roy Schulte, Gartner Inc.

5 Herausforderungen für Big Data

- Eine Geschäftsfallidentifizierung mit zweifelsfreier Big-Data-Rechtfertigung fällt schwer, da Qualität und Nutzen der Analyseergebnisse oft kaum absehbar sind
- Die Grenze zwischen Big-Data und traditionellen BI-/DW-Szenarien ist fließend
- Eine Big-Data-Strategie kann Teil der BI-Strategie sein oder komplett neu entwickelt werden, falls sich neue Geschäftsfelder um die Daten aufbauen
- Zur Analyse poly-strukturierter Daten muss ein Datenaustausch mit herkömmlichen Analysen konzeptionell, architektonisch und technisch realisiert werden
- Es fehlt an geeigneten Big-Data-Entwicklern und –Analysten, für die sich im englischen der Begriff „Data Scientist“ herausbildet, zur Analyse und Visualisierung der Daten
- Insbesondere in Deutschland müssen die Implikationen durch Datenschutz und Ethik berücksichtigt werden, z.B. bei der Auswertung sozialer Netzwerke
- Skalierbarkeit, Performance, Realtime-Bereitstellung und Wartung müssen trotz mangelnder Reife der Produkte sichergestellt werden

[vgl. Barc, in: Computerwoche 09/12]

5 Analytics der neuen Generation



Wachstum in drei Dimensionen

[IDC 2012]

BMBF richtet Big Data Kompetenzzentrum unter Federführung der TU Berlin ein



Prof. Dr. Volker Markl mit Bundesministerin Prof. Dr. Johanna Wanka

Das Bundesministerium für Bildung und Forschung (BMBF) wird die Forschung im Bereich Big Data und IT-Sicherheit in Deutschland neu aufstellen. Dazu werden in diesem Jahr zwei Kompetenzzentren eingerichtet. Eines davon wird das Berlin Big Data Center (**BBDC**) sein, das unter Federführung der TU Berlin steht. Ein weiteres ist das Competence Center for Scalable Data Services and Solutions (**ScaDS**), das unter der Leitung der TU Dresden steht. Beide Vorhaben werden mit insgesamt rund 10 Millionen Euro unterstützt. Den Startschuss zu den neuen Kompetenzzentren gab Johanna Wanka, Bundesministerin für Bildung und Forschung im Rahmen der CeBIT am 10. März 2014 in Hannover.

Danke Roland !



5 Literatur

- *Bange, Carsten/Grosser Timm* (2012): Daten-Management: Big Data – BI der nächsten Generation. URL: <http://www.computerwoche.de/software/bi-ecm/2505617/>. Abruf: 17.04.2012.
- *BARC* (2013): Big Data Survey Europe, BARC-Institut, Würzburg
- *BITKOM* (2013): Big Data im Praxiseinsatz – Szenarien, Beispiele, Effekte
- *Eckerson, Wayne* (2011): Big Data Analytics: Profiling the Use of Analytical Platforms in User Organizations. In: *TDWI White Papers*.
- *Franks, Bill* (2012): Taming the big data tidal wave. Hoboken, New Jersey: John Wiley & Sons Inc.
- *Gansor et al.* (2010): Von der Strategie zum Business Intelligence Competency Center (BICC), München: Hanser
- *Gantz, John/Reinsel David* (2011): 2011 Digital Universe Study: Extracting Value from Chaos.
- *IBM Global Business Services* (2012): Analytics: Big Data in der Praxis, IBM Institute for Business Value
- *Klein, Dominik/Tran-Gia, Phuoc/Hartmann, Matthias* (2013): Big Data. In: Informatik Spektrum 36 (3), S. 319-323.
- *LaValle, Steve et al.* (2010): Big data, analytics and the path from insights to value. In: *MIT sloan management review* 52, S. 21–32.
- *Proffitt, Brian* (2012): Big data tools and vendors. URL: <http://www.itworld.com/big-datahadoop/251912/big-data-tools-and-vendors?page=0,0>, Abruf am 24.04.2012.
- *Russom, Philip* (2011): Big Data Analytics, TDWI Best Practices Report
- *Storage Consortium*: Big Data und die Anforderungen aus Infrastruktursicht. URL: <http://www.storageconsortium.de/content/node/1099>, Abruf am 24.04.2012.
- *Wartala, Ramon* (2012): Hadoop. München: Open Source Press.
- *White, Tom* (2011): Hadoop, 2. Aufl. Beijing [u.a.]: O'Reilly.
- *Zacher, Matthias* (2012): Big Data Analytics in Deutschland 2012. In: *SAS IDC White Paper*
- *Zikopoulos, Paul* (2012): Understanding big data. New York: McGraw-Hill.