



Manufacturing Analytics

Internet of Things / Industrial Internet

Lehrstuhl für Wirtschaftsinformatik,
insb. Prozesse und Systeme
Universität Potsdam

Chair of Business Information Systems,
esp. processes and systems
University of Potsdam

Univ.-Prof. Dr.-Ing. habil. Norbert Gronau
Lehrstuhlinhaber | Chairholder

August-Bebel-Str. 89 | 14482 Potsdam | Germany

Tel +49 331 977 3322

Fax +49 331 977 3406

E-Mail ngronau@lswi.de

Web lswi.de

Manufacturing Analytics

Lernziel

- Die Rolle von Daten, Informationen und Wissen in Systemen des „Data Warehousing“, „Data Mining“, „Decision Support Systems“, „Wissensmanagement“, „Business Intelligence“ und „Business Analytics“ beschreiben können
- Bedingungen für einen optimalen Trade-off der Datenveredelung kennen und an einem praktischen Beispiel erläutern können
- Charakteristiken, Vor- und Nachteile bei den Verarbeitungsstrategien „Distributed Memory“, „Parallelisierbare Verarbeitung“ und „MapReduce“ kennen
- Komponenten der MapReduce-Arbeitsumgebung kennen und Zusammenspiel bei der Verarbeitung verstehen
- Herausforderungen im Manufacturing Analytics und Begründung zur Verortung im Prozessverlauf

Einstiegsvideo

Anwendungsfall PROXIA MES





Business Intelligence

Business Analytics

Umgang mit Big Data

Anwendungsvoraussetzungen

Begriffe

Business Intelligence

- Zur Gewinnung von Erkenntnissen mittels statistischer und quantitativer Analysen
- Fokussiert aktuelle Probleme auf Daten der Vergangenheit
- Ermöglichung besserer Entscheidungen im Operativen und/ oder Strategischen

Business Analytics (auch Advanced Analytics)

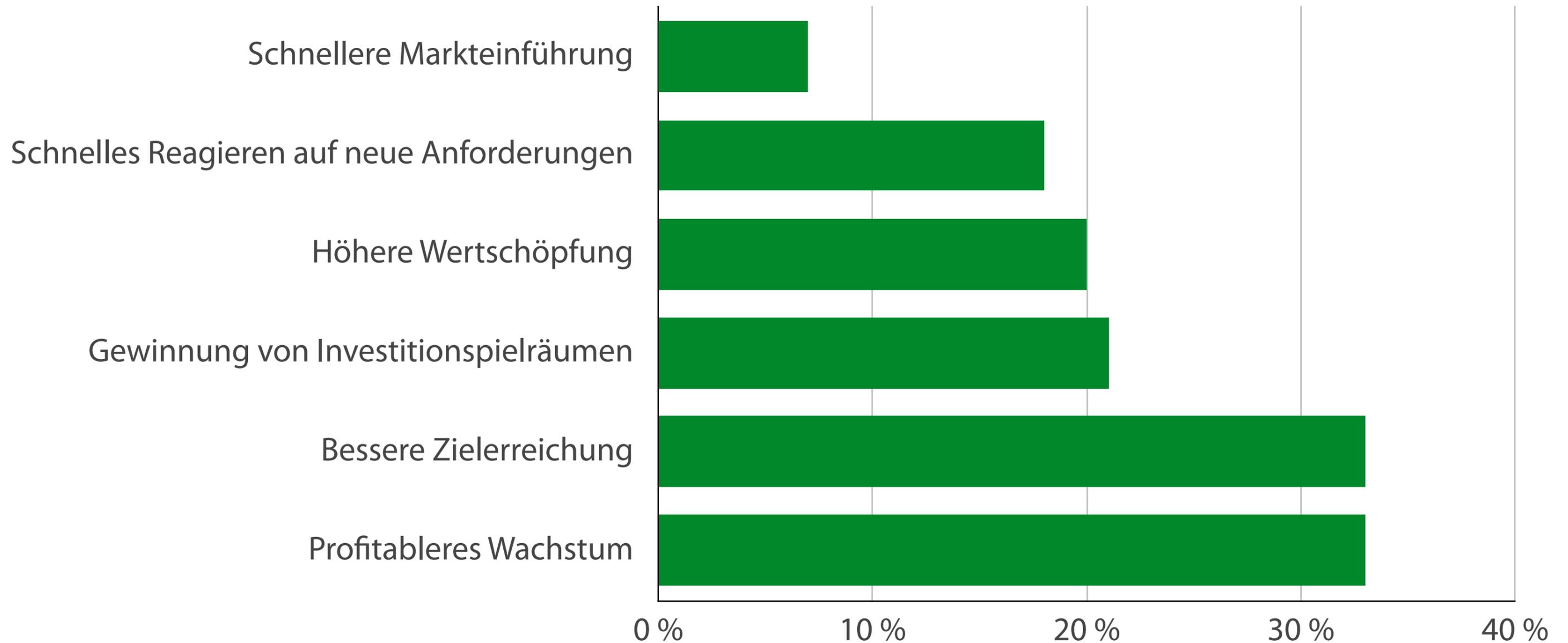
- Erweitert den Bereich des BI
- Berücksichtigung von Erklärungs- und Vorhersagemodellen
- Fokussiert die Prognose zukünftiger Probleme und Lösungen

Manufacturing Analytics - Production Analytics

- Nutzung von Daten, BI- und BA-Techniken im Produktionsbereich
- Zur Vorbereitung von Entscheidungen und Handlungen
- Integration mit einem faktenbasierten Produktionsmanagement

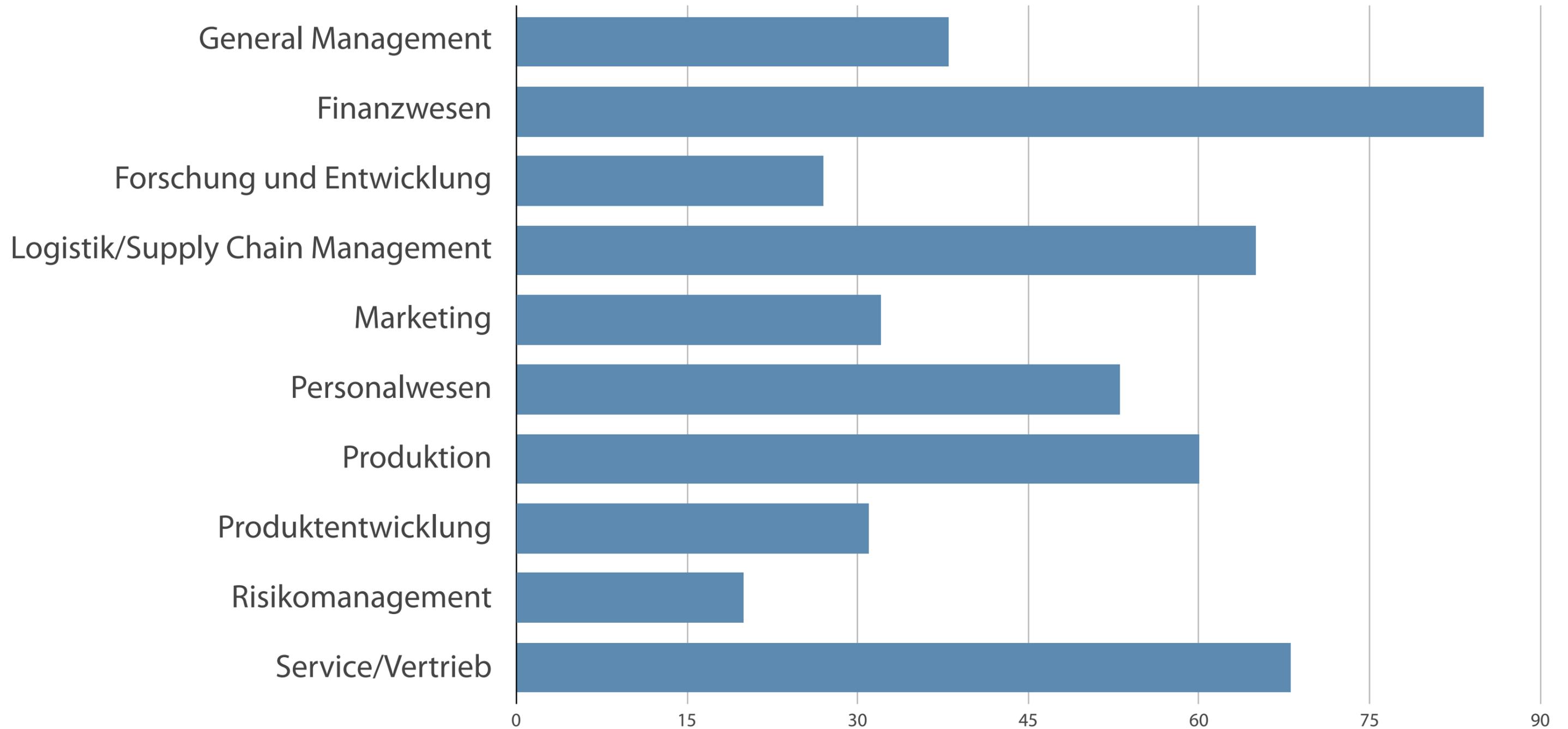
Als Beispiel dient die Erschließung bisher nicht genutzter Daten einer Produktionsanwendung

Potenziale von Business Analytics für den Erhalt der Wettbewerbsfähigkeit



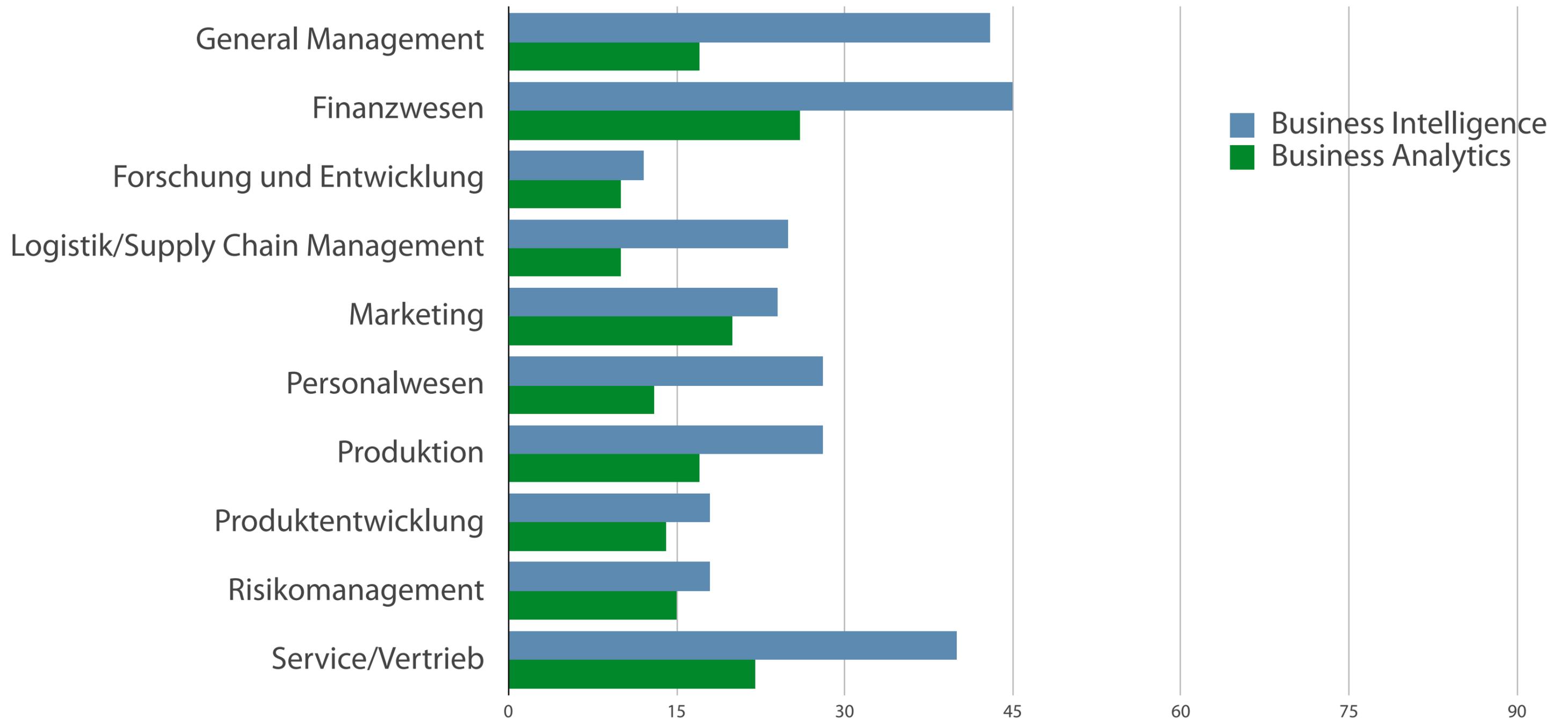
Die gezeigten Potenziale können in der Industrie bislang kaum erschlossen werden.

Datennutzung in unterschiedlichen Unternehmensfunktionen



Datennutzung erfolgt am intensivsten im Finanzwesen.

Anwendungsvergleich in unterschiedlichen Unternehmensfunktionen



Bzgl. der Nutzung von BI und BA in den Fachabteilungen liegt die Fertigung im Mittelfeld.



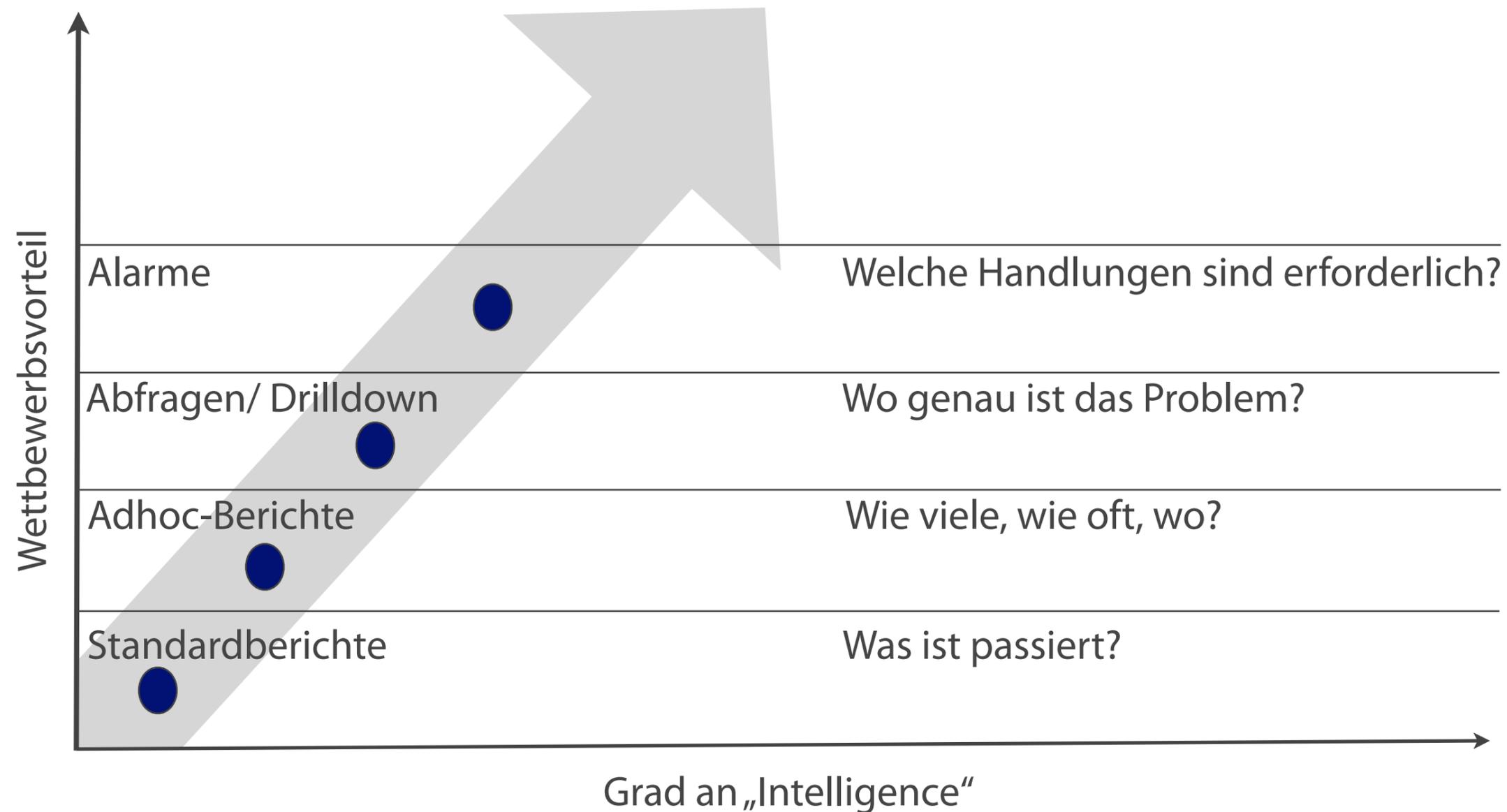
Business Intelligence

Business Analytics

Umgang mit Big Data

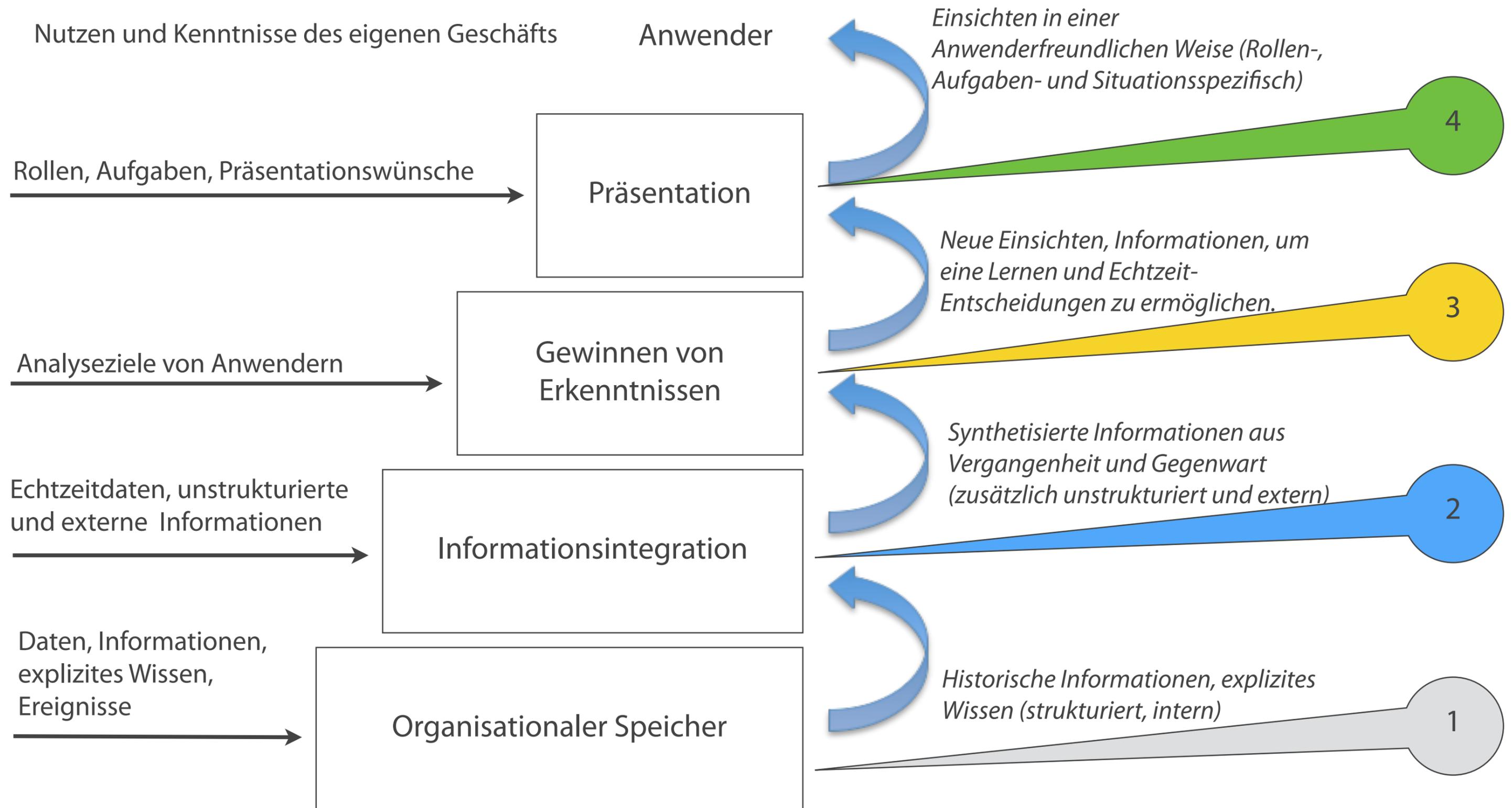
Anwendungsvoraussetzungen

Möglichkeiten durch BI (nach Davenport)

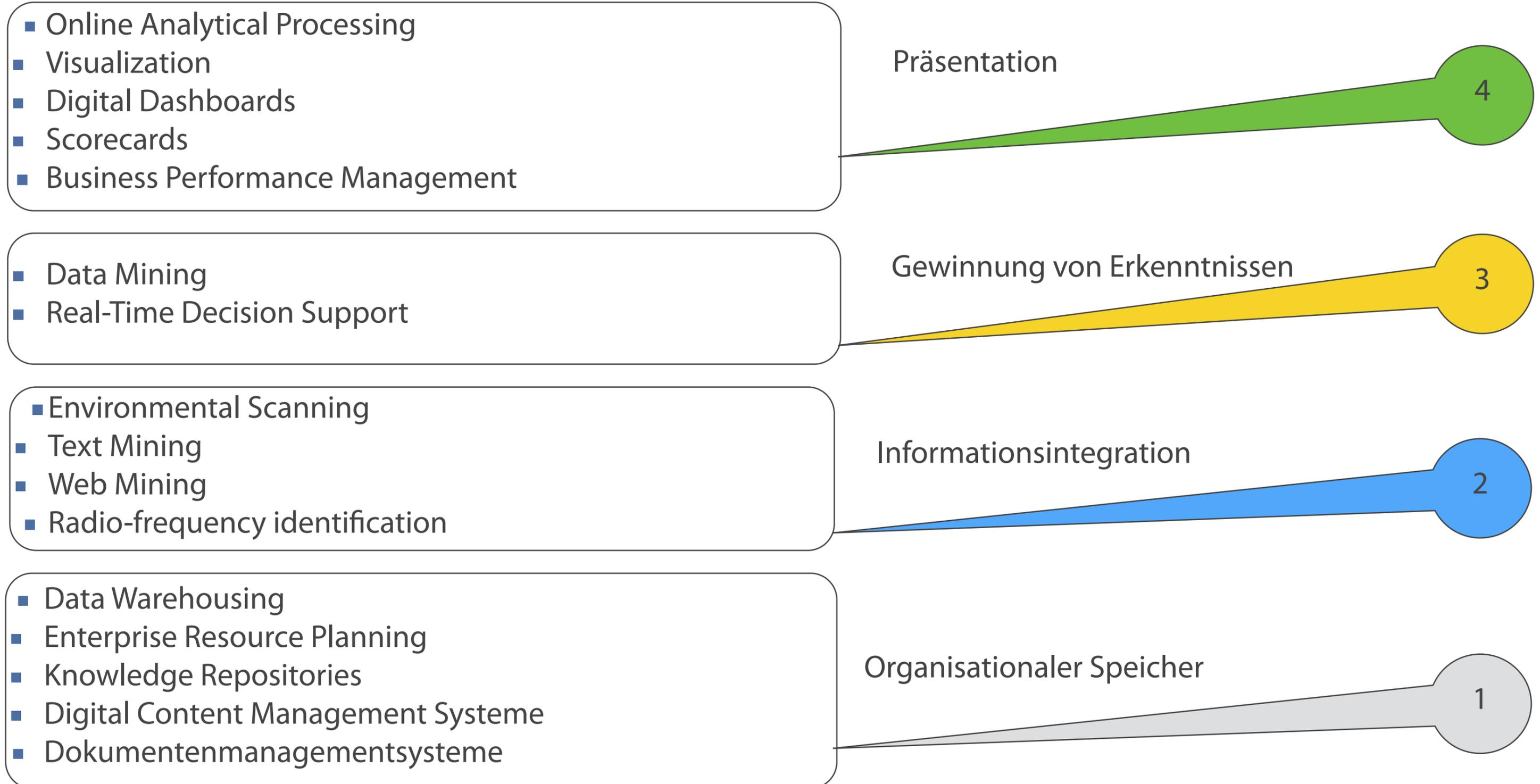


Gestaltungs-, Optimierungs- und Innovationspotenziale können nicht allein durch die Anwendung klassischer, überwiegend vergangenheitsorientierter Reportingsysteme (Business Intelligence) erschlossen werden.

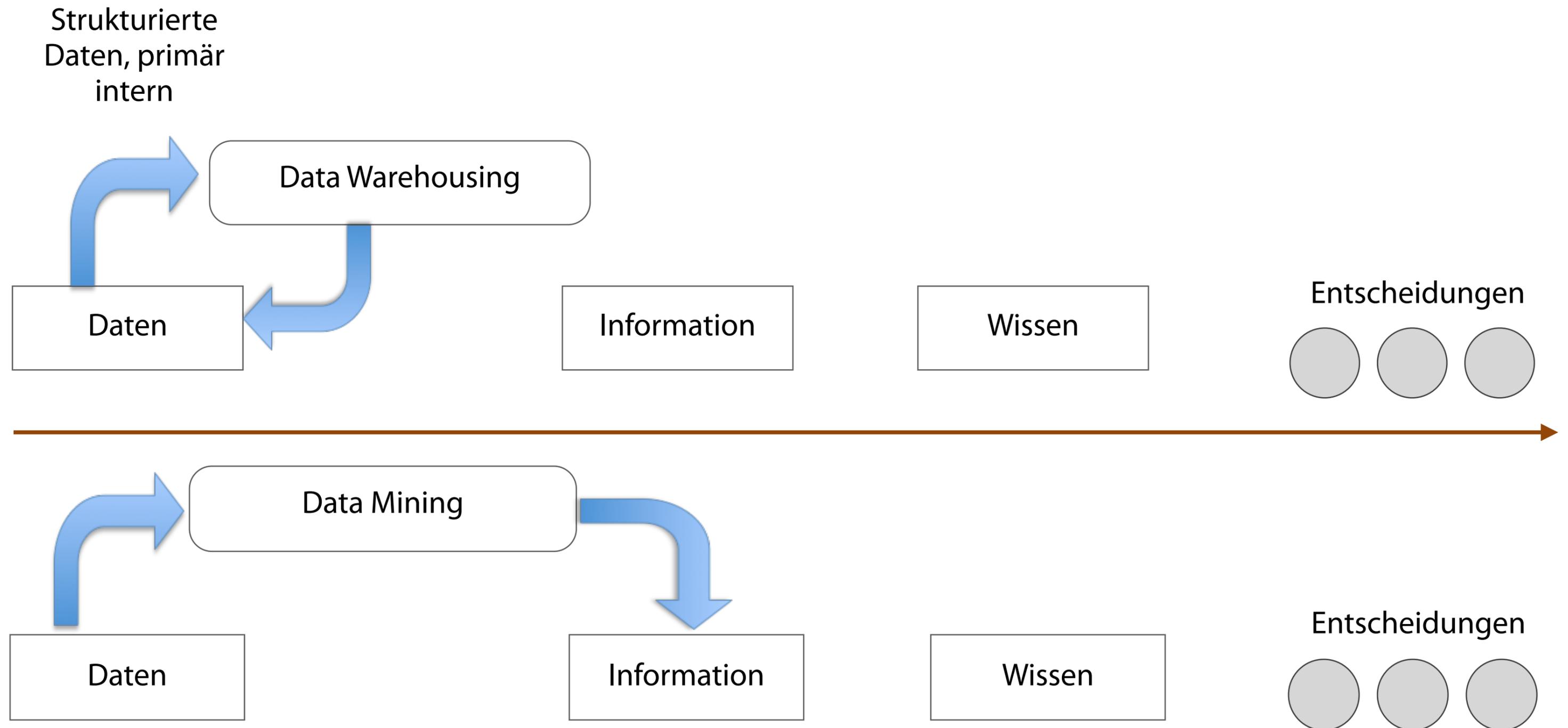
Fähigkeiten, die für Business Intelligence benötigt werden



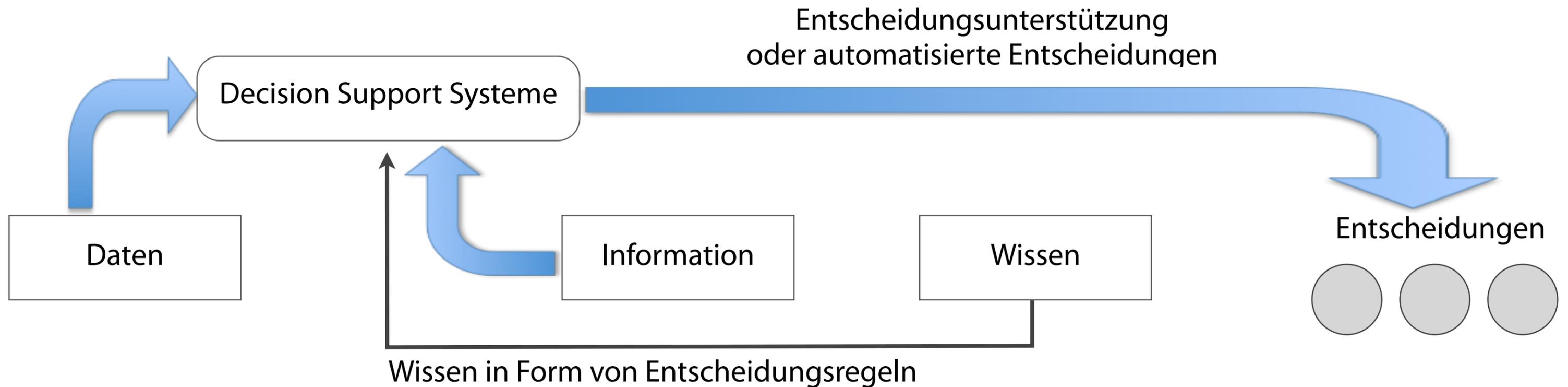
Techniken des BI im Überblick



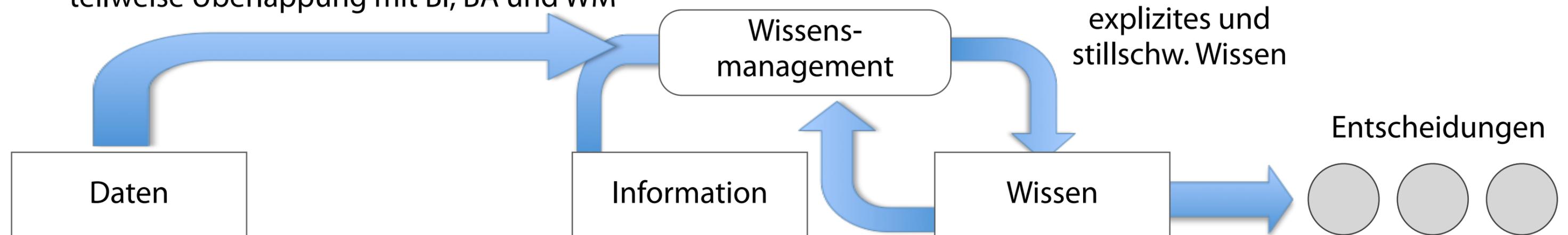
Rolle von Daten, Informationen und Wissen bei Data Warehousing und Data Mining



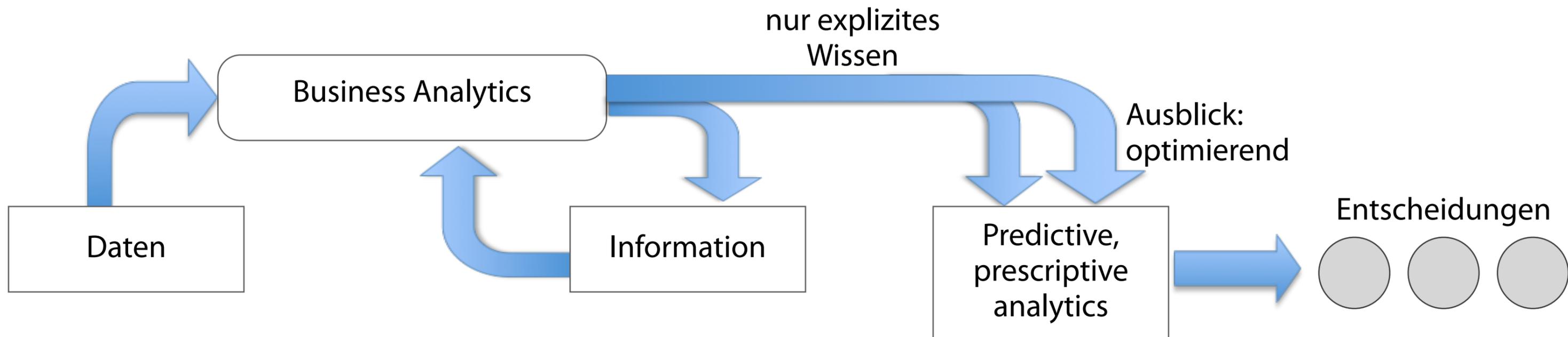
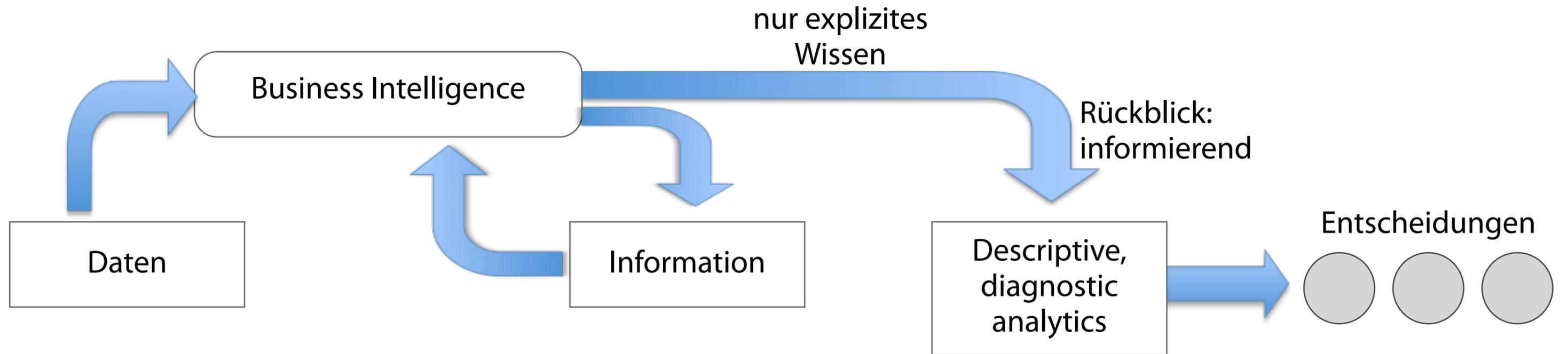
Rolle von Daten, Informationen und Wissen bei DSS und Wissensmanagement



Wissensidentifizierung in Daten und Informationen;
teilweise Überlappung mit BI, BA und WM



Rolle von Daten, Informationen bei Business Intelligence und Business Analytics

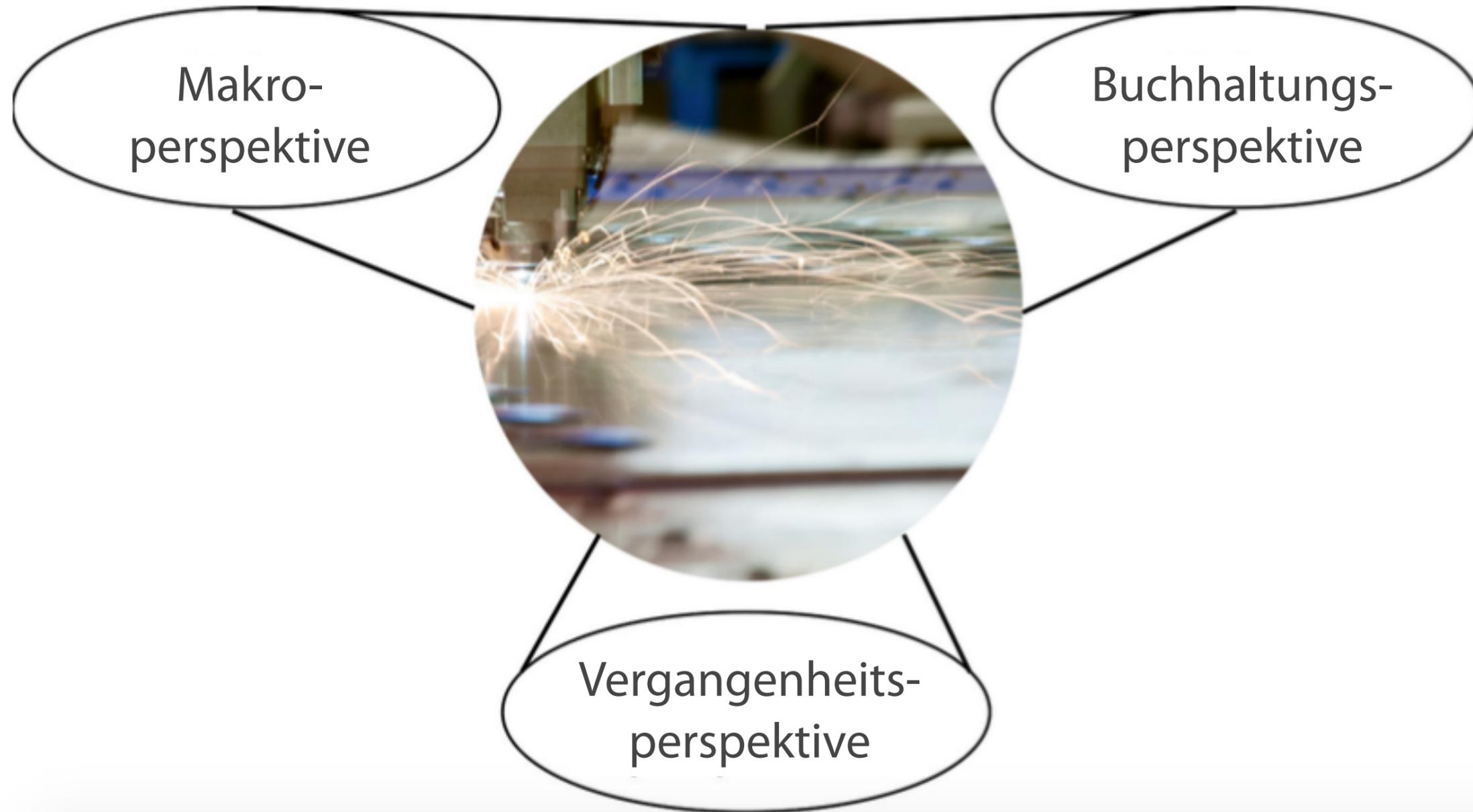


Entscheidungen im Produktionsmanagement

- Einbindung entscheidungsbefugter Manager ins Tagesgeschäft
 - Entscheidungsfindung auf Basis unvollständiger Daten
 - Insellösungen und separate Datengrundlagen
 - Treffen unsystematischer Entscheidungen
- Häufige
Entscheidungsfehler im
Produktionsmanagement**
- Schwacher Zusammenhang aus Daten, Informationen, Wissen und Schlussfolgerungen
 - Häufiges Variieren von Entscheidungsabläufen
 - Intransparente Entscheidungsverantwortlichkeiten
 - Ignorieren von Lernerkenntnissen aus Fehlentscheidungen

Häufige Entscheidungsprobleme im Produktionsmanagement sind primär auf bereichsisolierte Betrachtungen zurückzuführen.

Vorherrschende Perspektiven in der Fertigung



Eine angemessene Perspektive auf die Fertigung integriert mehrdimensionale Auswertungen und zieht heterogene Fragestellungen in Betracht.



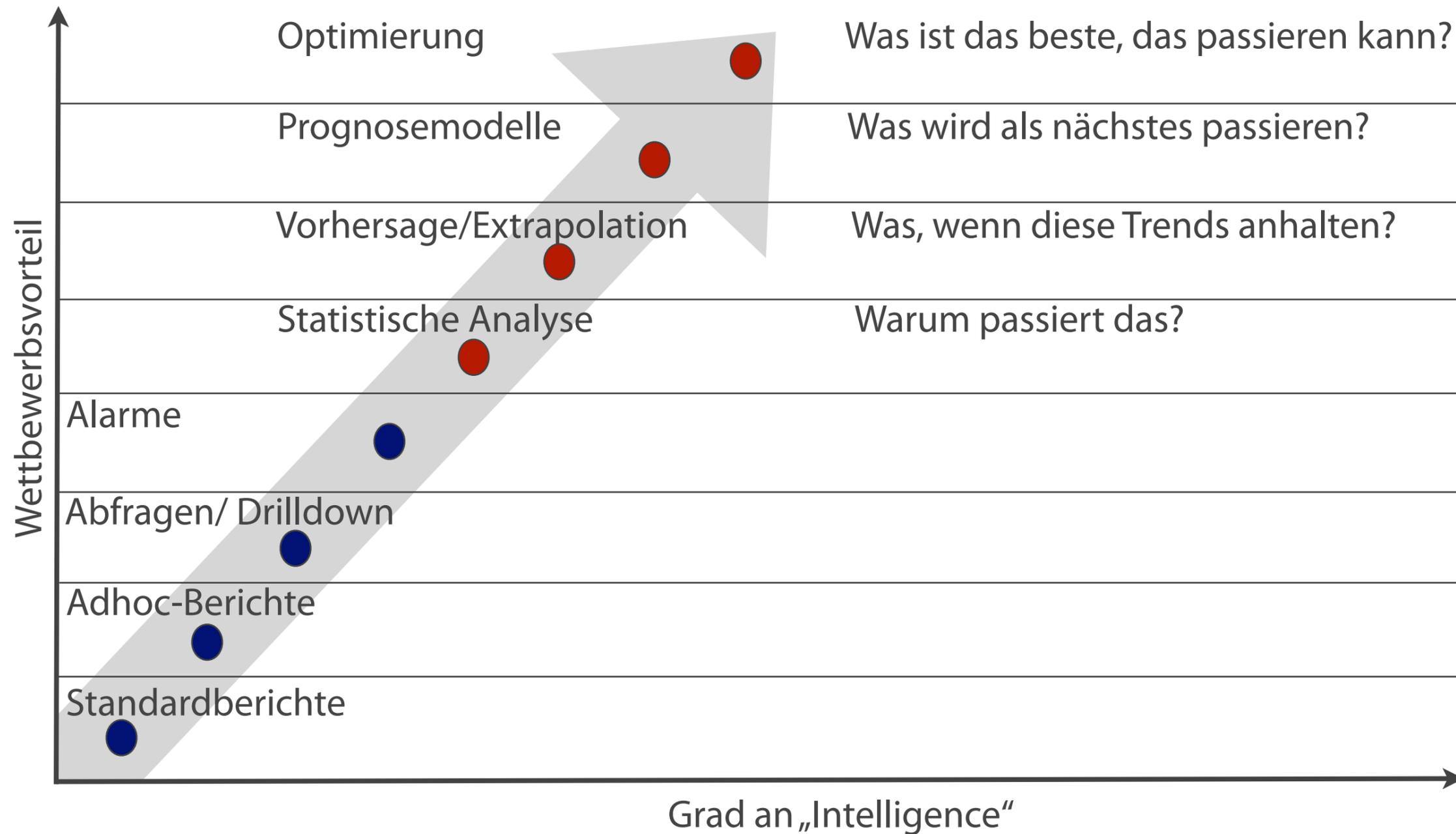
Business Intelligence

Business Analytics

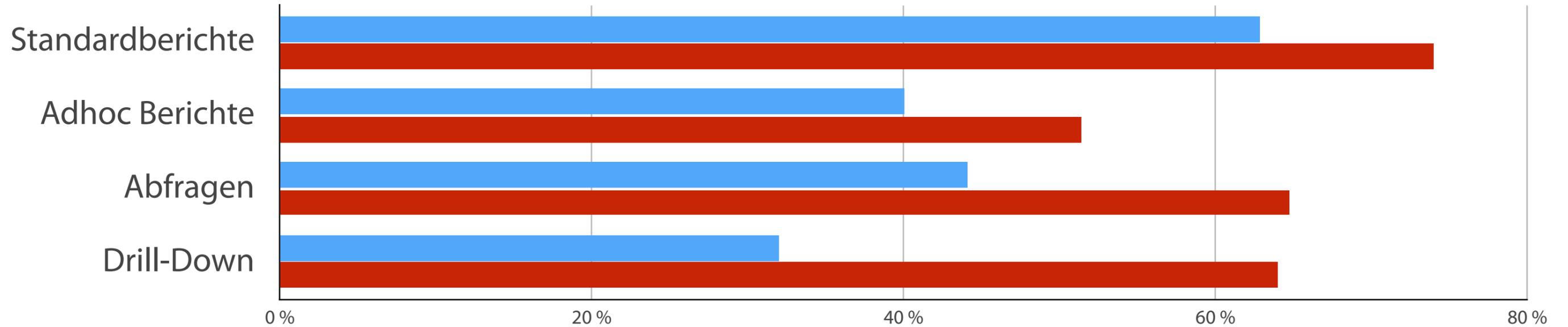
Umgang mit Big Data

Anwendungsvoraussetzungen

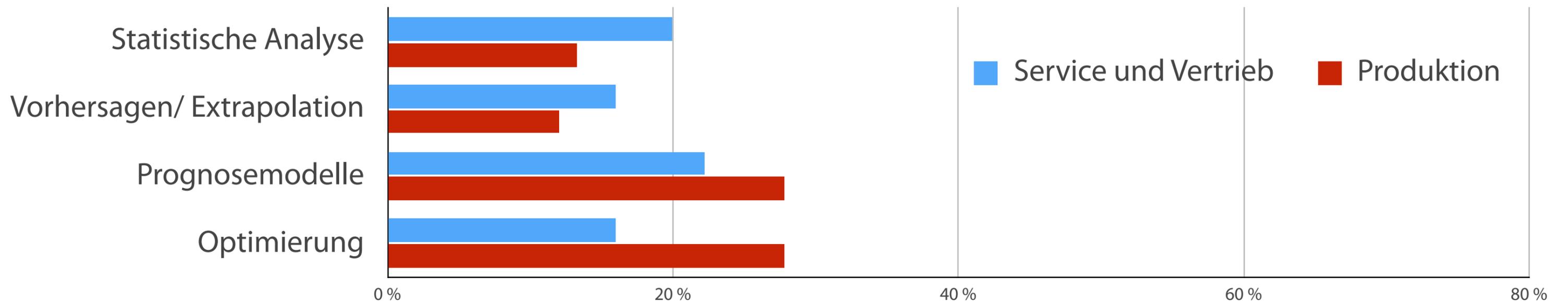
Möglichkeiten durch BA (nach Davenport)



Nutzung von BI und BA



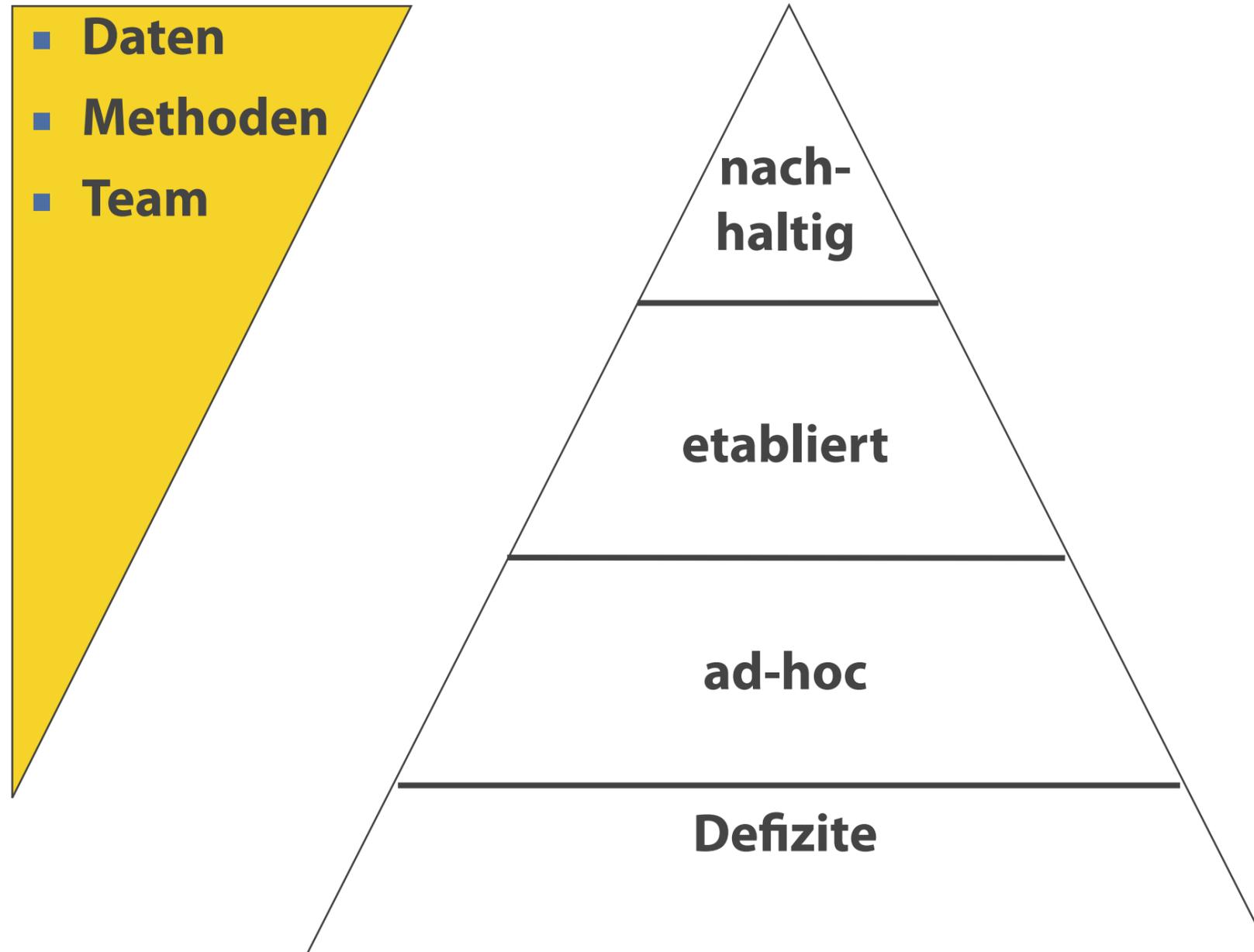
Verwendung analytischer Methoden



Während BI-Funktionen in der deutschen Industrie bereits umfassend verwendet werden, hinken BA-Funktionen bei der Nutzung von großen Datenmengen zu Prognose, Simulation und Optimierung hinterher.

Reifegradstufen

Den Unternehmenswert im Fokus



Eine nachhaltige Steigerung der Wettbewerbsfähigkeit ist dann möglich, wenn ein Dreiklang aus Daten, Methoden und im Team verfügbaren Methodenwissen besteht.



Business Intelligence

Business Analytics

Umgang mit Big Data

Anwendungsvoraussetzungen

„Big Data“

- Sammlung von strukturierten und unstrukturierten Daten
- Mitunter inkl. Analyse und Auswertung
- Daten mit komplexen V-Charakteristiken

„Big Data“-Umwelten

- berücksichtigen eine verteilte Verarbeitung
- hohe Abstraktion sowie
- offene Programmierschnittstellen

Charakteristiken von „Big Data“-Datenmengen

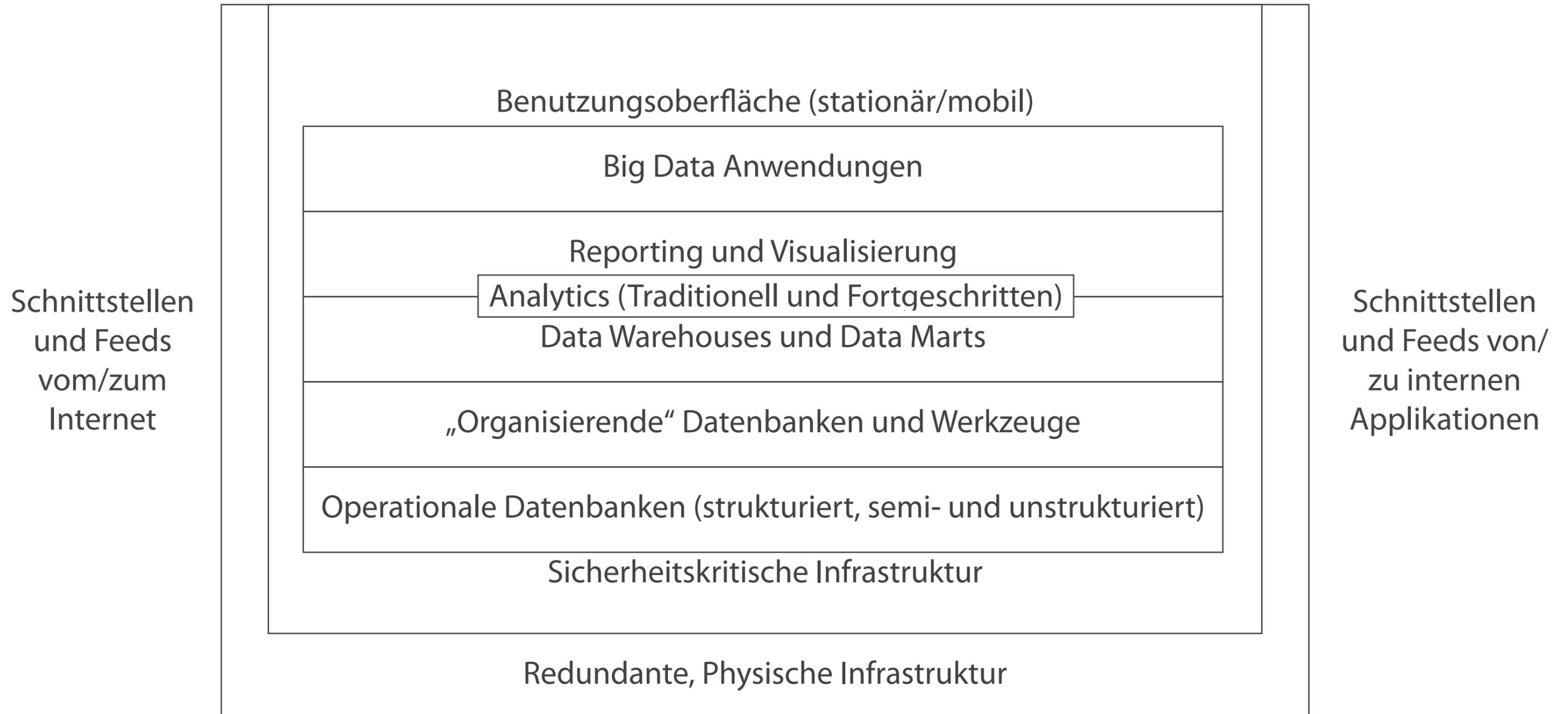
- Volume (Umfang)
- Variety (Bandbreite)
- Velocity (Geschwindigkeit)
- Veracity (Wahrhaftigkeit)

„Big Data“-Lösungen

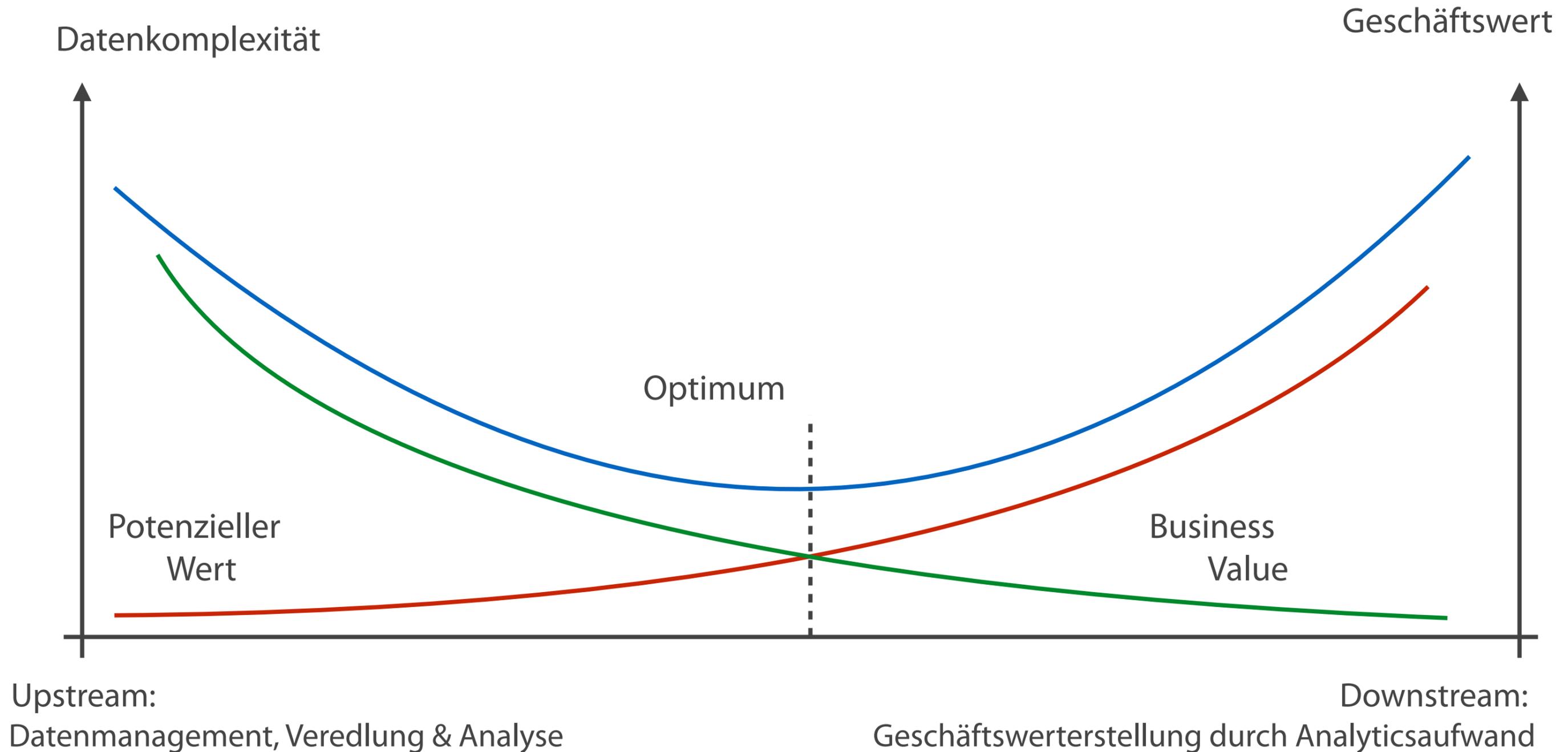
- berücksichtigen die physische Infrastruktur,
- Sicherheitsfragen und
- Analysewerkzeuge

Da klassische Ansätze der Datenverarbeitung mit den Anforderungen von „Big Data“ nicht umgehen können, werden hier neue Ansätze gefordert.

„Big Data“-Architektur



Trade-off der Datenveredelung



Eine wertoptimale Veredelung von Daten befindet sich im Fit von Geschäftswertgenerierung und Potenzialverlust.

Ansätze zum Umgang mit Big Data

Datencharakteristik

- Datengrößen
- Datentypen
- Datenspeicherort

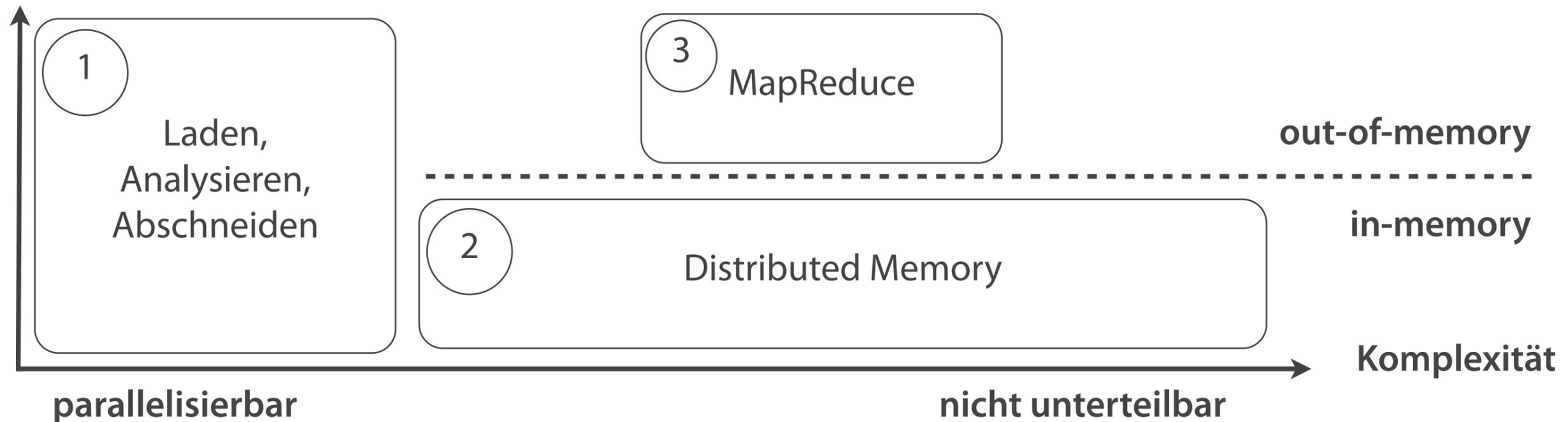
Rechenplattform

- Desktop-Maschine
- Mehrere Maschinen
- Cluster-Maschine

Analysecharakteristik

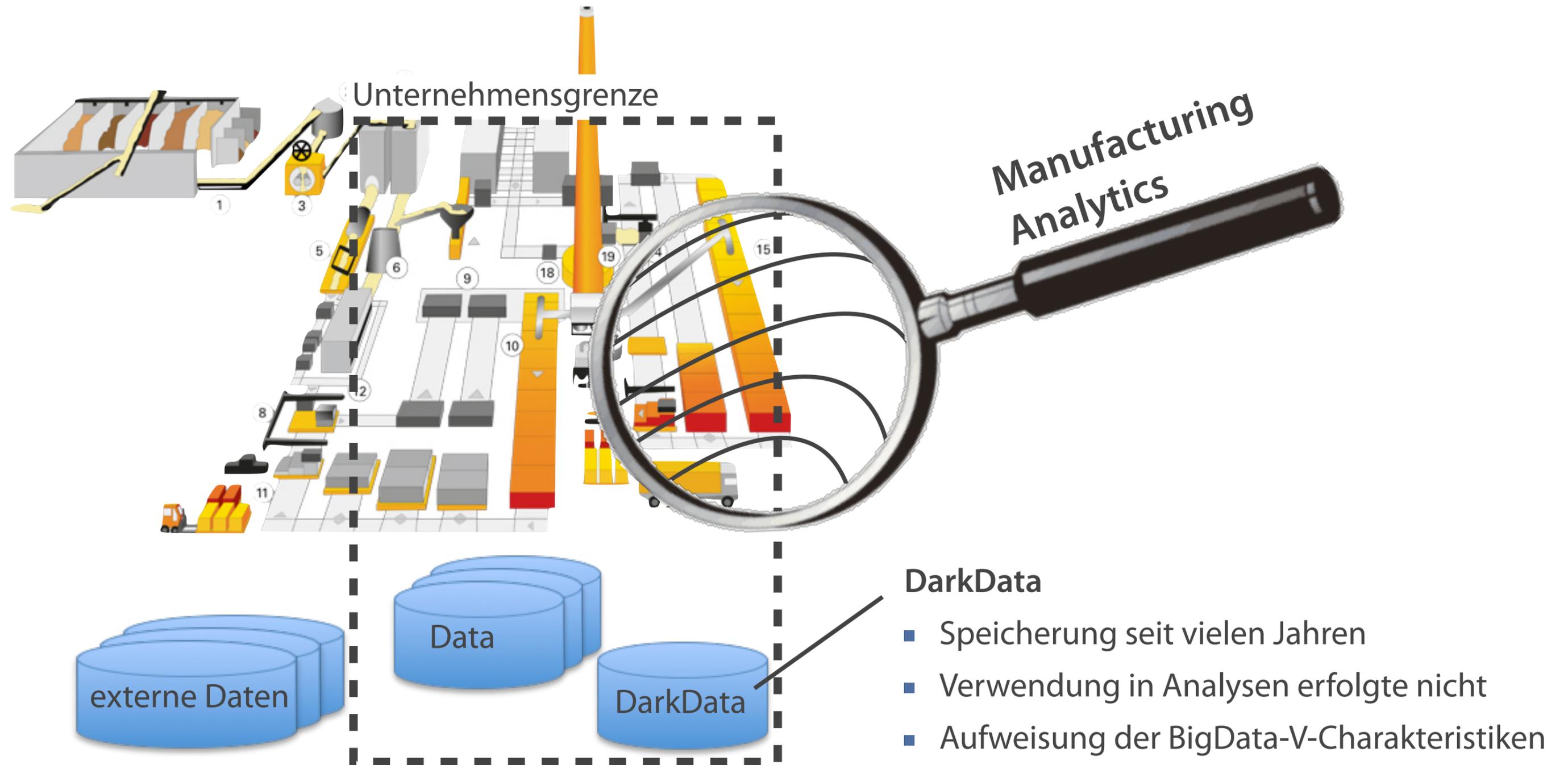
- Parallelisierbarkeit
- Ressourcenbedarf

Ressourcen- anforderungen



Die Auswahl des „Big Data“-Ansatzes ist abhängig von der Komplexität der Daten und den Ressourcenanforderungen für die Analyse.

Ein Beispiel des Manufacturing Analytics für die Ansätze zum Umgang mit Big Data



Als Beispiel dient die Erschließung von DarkData einer Produktionsanwendung.

Datencharakteristik

- Beliebige Datenformate, solange sie geteilt werden können
- Zentraler Datenspeicherort

Rechenplattform

- Desktop-Maschine
- Cluster-Maschine

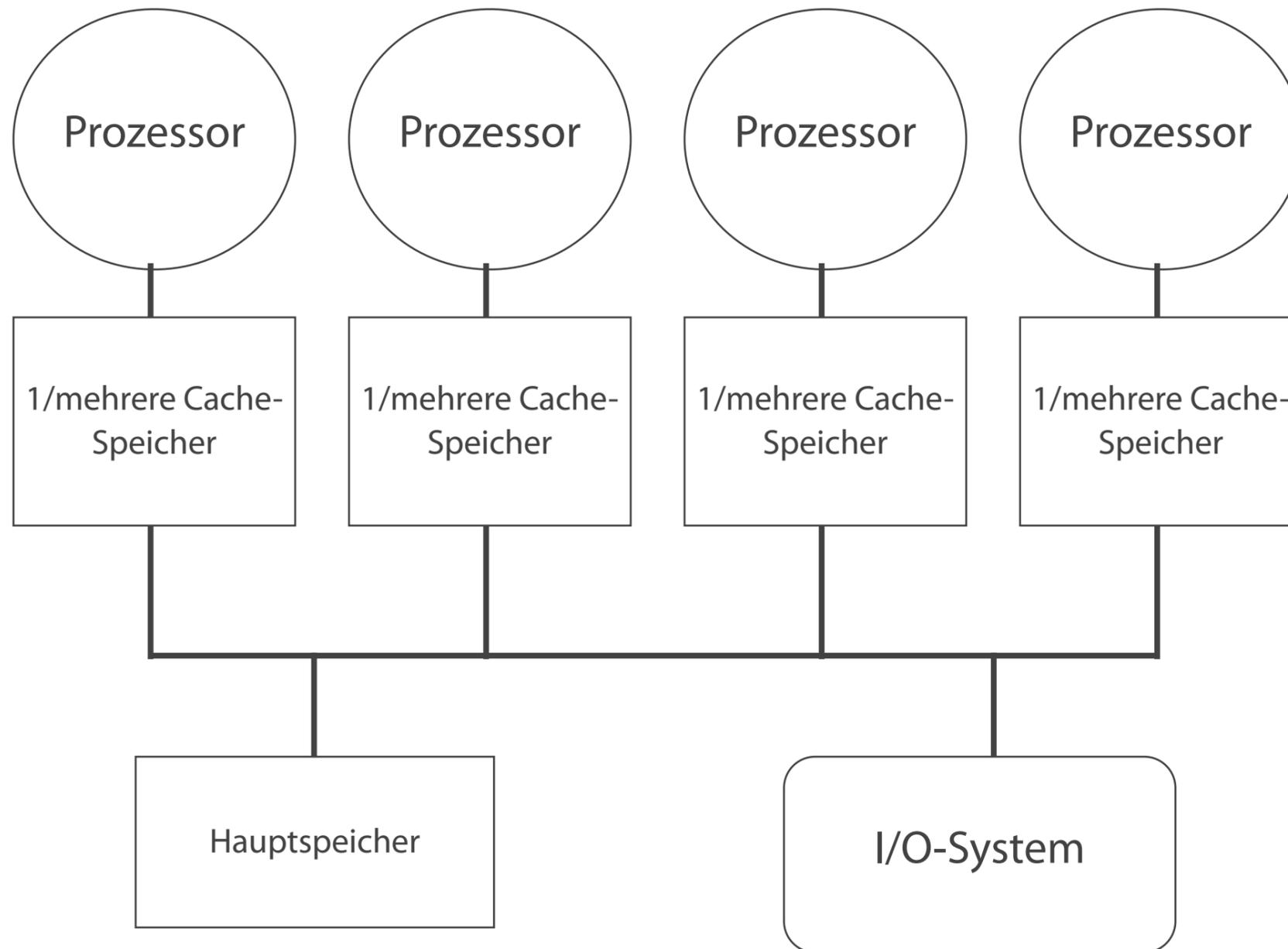
Analysecharakteristik

- Jede Iteration einer Schleife muss parallelisierbar sein
- Daten jeder Iteration müssen den Ressourcen und RAM der Maschine genügen

„Uniform Memory“-Ansatz

- Ein Multiprozessoren System mit nur einem globalen Speicher.
- Operationen verteilter Prozessoren greifen auf diesen Speicherbereich zu.
- Da Bandbreite und Latenzzeit in Einklang sein sollten, werden sie auch als SMPs bezeichnet.

Engpässe in der Verfügbarkeit von Rechenleistung können mit Hilfe verteilter Verarbeitungsansätze gelöst werden.



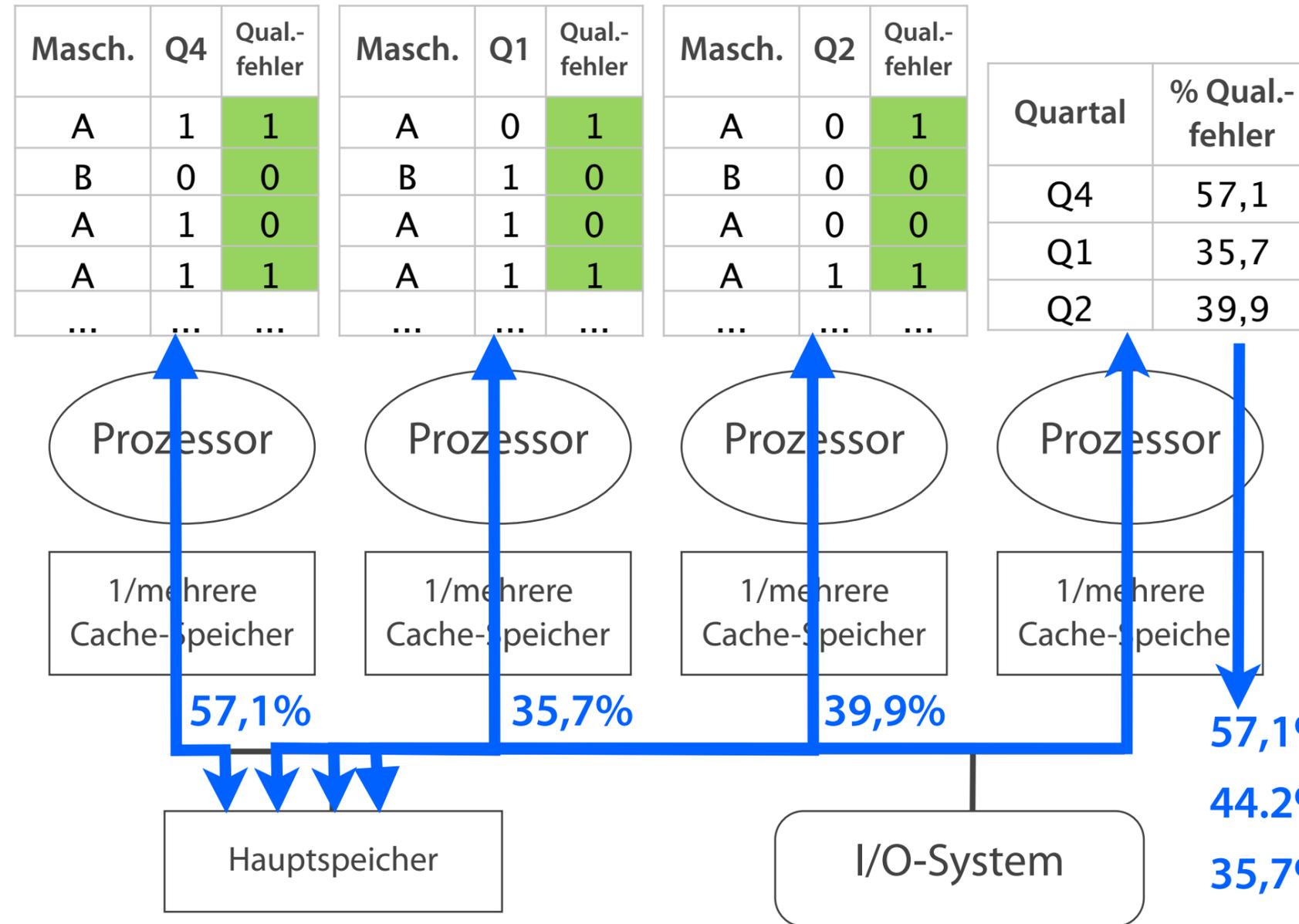
Das Augenmerk in Systemen mit globalen Speicher ist auf eine sinnvolle Verteilung der Berechnungen zu legen, damit die technikhärenten Potenziale erschlossen werden können.

Parallelisierbare Verarbeitungen

Bsp: Qualitätsfehleranalyse

Datensammlung

Maschine	Q4	Q1	Q2	Qual.-fehler
A	1	0	0	1
B	0	1	0	0
A	1	1	0	0
A	1	1	1	1
B	0	0	0	1
A	0	0	0	1
A	0	1	1	0
A	1	0	1	0
C	0	0	1	0
B	1	1	0	1
C	0	1	0	0
A	1	0	0	1
A	0	0	1	1
D	1	0	0	0



Das Augenmerk in Systemen mit globalen Speicher ist auf eine sinnvolle Verteilung der Berechnungen zu legen, damit die technikhärenten Potenziale erschlossen werden können.

Datencharakteristik

- Daten müssen in den kollektiven Speicher verfügbarer Maschinen passen
- Dezentrale Datenspeicherorte

Rechenplattform

- Desktop-Maschine (Prototypen, Teilmenge an Daten)
- Cluster-Maschine oder Cloud

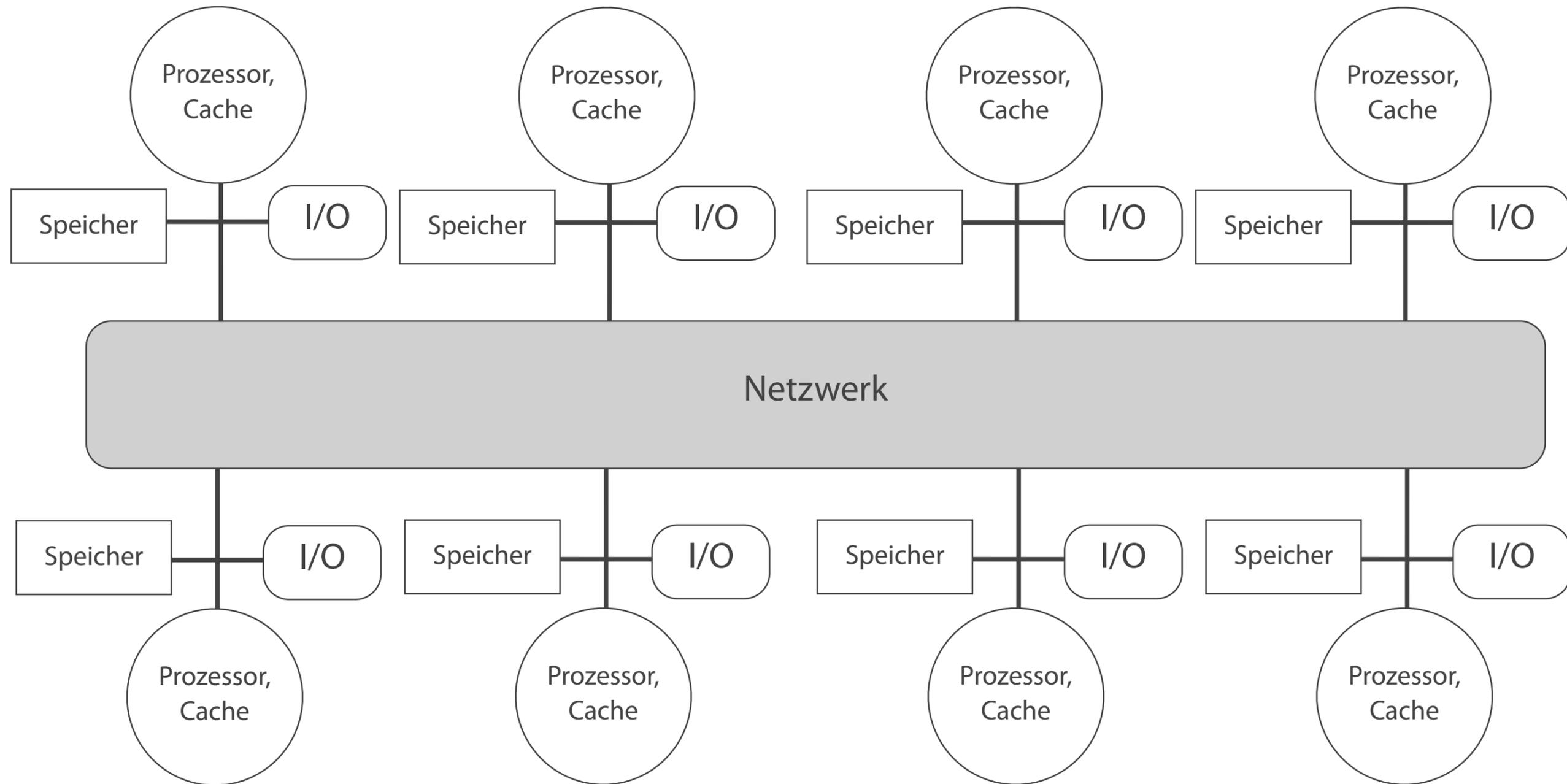
Analysecharakteristik

- müssen z.T. parallelisierbar sein und mit unterstützten Funktionen und Datenstrukturen berechnet werden können
- Teile müssen dem Ressourcenbedarf und RAM der Recheneinheit genügen

„Distributed Memory“-Ansatz

- Multiprozessoren System
- Jede Recheneinheit mit eigenem, privatem Speicher
- Operationenausführung auf eigenem Speicherbereich
- Übertragung entfernter Daten über Kommunikationsvorgänge

Engpässe in der Datenverfügbarkeit können mit Hilfe verteilter Speicheransätze gelöst werden.



Das Augenmerk in Systemen mit verteiltem Speicher ist auf eine sinnvolle Verteilung der Daten zu legen, damit technikhärenten Potenziale erschlossen werden können.

Zur Ausführung einer einzelnen Aufgabe:

Algorithm *Task_t*:

```
Do some computation;  
send one message on each channel of a (possibly empty) subset of Outt;  
repeat  
    receive message on  $c_1 \in I_{n_t}$  and  $B_1 \rightarrow$   
        Do some computation;  
        send one message on each channel of a (possibly empty) subset of Outt;  
    or...  
    or  
        receive message on  $c_{n_t} \in I_{n_t}$  and  $B_{n_t} \rightarrow$   
            Do some computation;  
            send one message on each channel of a (possibly empty) subset of Outt;  
until global termination is known to t.
```

Die Ausgestaltung von verteilt arbeitenden Algorithmen erfordert aufwändige Kommunikationen zwischen den Recheneinheiten und es muss gewährleistet werden, dass die Vorteile des Ansatzes realisiert werden können.

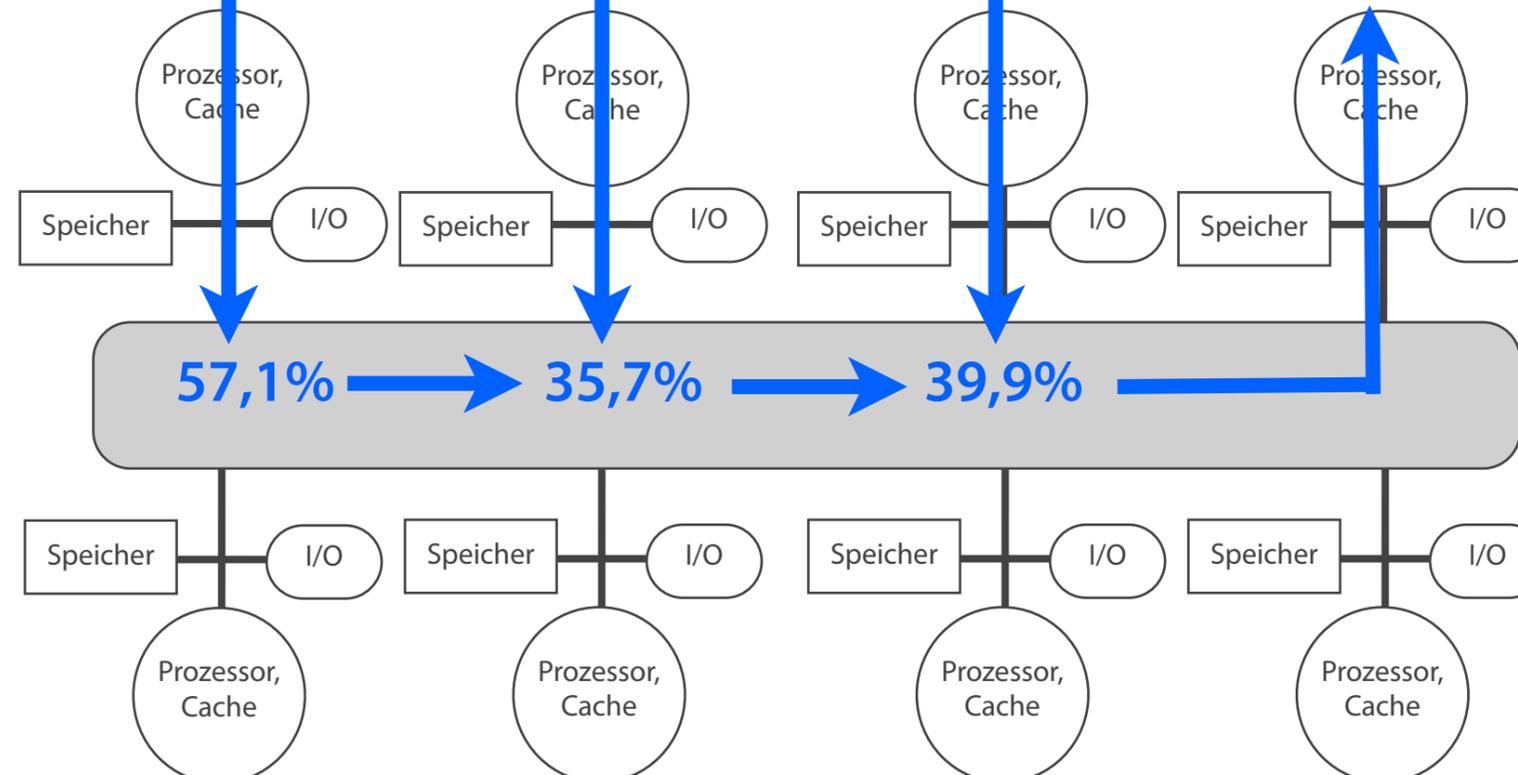
Distributed Memory

Bsp: Qualitätsfehleranalyse

Datensammlung

Maschine	Q4	Q1	Q2	Qual.-fehler
A	1	0	0	1
B	0	1	0	0
A	1	1	0	0
A	1	1	1	1
B	0	0	0	1
A	0	0	0	1
A	0	1	1	0
A	1	0	1	0
C	0	0	1	0
B	1	1	0	1
C	0	1	0	0
A	1	0	0	1
A	0	0	1	1
D	1	0	0	0

Masch.	Q4	Qual.-fehler	Masch.	Q1	Qual.-fehler	Masch.	Q2	Qual.-fehler	Quartal	% Qual.-fehler
A	1	1	A	0	1	A	0	1	Q4	57,1
B	0	0	B	1	0	B	0	0	Q1	35,7
A	1	0	A	1	0	A	0	0	Q2	39,9
A	1	1	A	1	1	A	1	1		
...		



57,1% max
 44,2% Durchschnl.
 35,7% min

Das Augenmerk in Systemen mit verteiltem Speicher ist auf eine sinnvolle Verteilung der Daten in Einklang mit den Algorithmen zu legen, damit technikhärenten Potenziale erschlossen werden können.

Datencharakteristik

- Riesige Datensatzgrößen
- Unstrukturierte Daten im HDFS
- Verteilte Datenspeicherorte

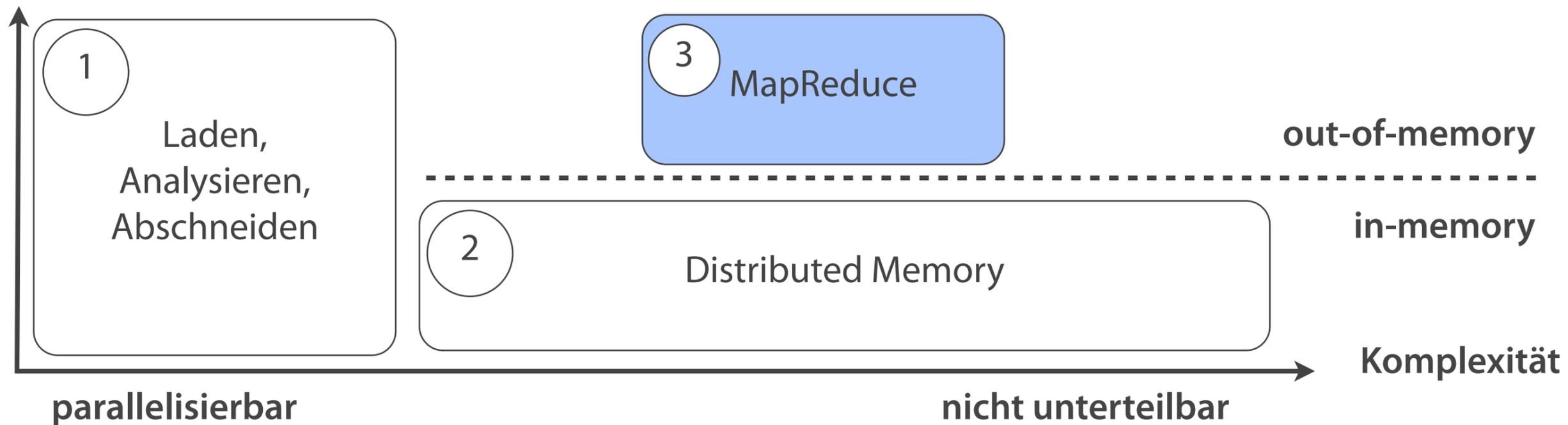
Rechenplattform

- Desktop-Maschine
- Mehrere Maschinen
- Cluster-Maschine

Analysecharakteristik

- müssen z.T. parallelisierbar sein
- erfordern immensen Ressourcenbedarf und passen nicht in den RAM

Ressourcenanforderungen



MapReduce-Ansatz

- Map: Operationen müssen auf einen handhabbaren Datenblock angewendet werden, sodass ein Zwischenergebnis erstellt wird.
- Reduce: Mehrere Zwischenergebnisse müssen zu einem Endergebnis konsolidiert werden

Anforderungen an eine MapReduce-Implementierung

- Jobverteilung entsprechend der Arbeitslast
- Jobmonitoring
- Fehlerrobustheit
- Nachbesserungsmöglichkeit
- Dezentrale Datenverteilung
- Dezentrale Datenverarbeitungsmöglichkeiten
- Bevorzugung kurzer Wege
- Zwischen- und Endergebnisbereitstellung
- Ergebnisbereitstellung

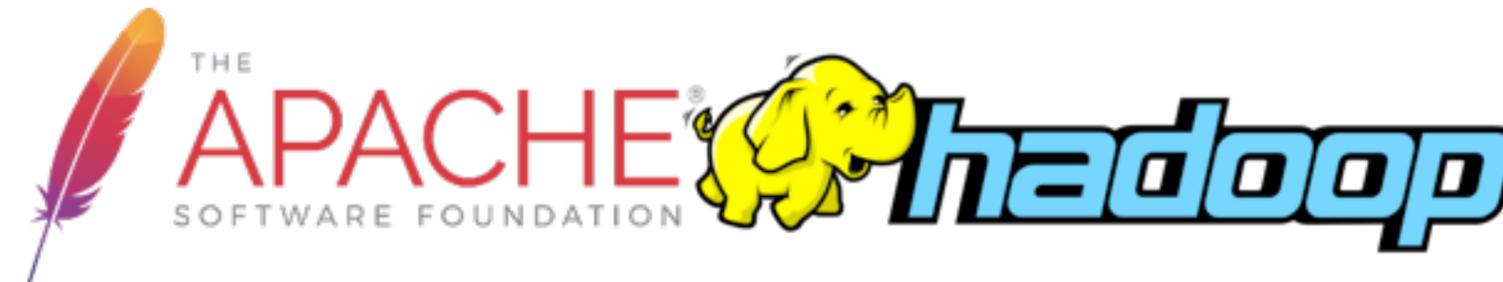
Aus dem scheinbar einfachen MapReduce-Datenverarbeitungsansatz lassen sich viele und sehr anspruchsvolle Implementierungsanforderungen ableiten.

- Pig
- Hive
- Mahout

- Zookeeper

Hadoop-Ecosystem:

Hadoop-Umwelt:



- Hadoop Distributed File System
- Verarbeitungsalgorithmen
- Code-Basis
- open-source-Community
- Markt für Umwelten

- HBase

- Chukwa

- ...

- Cassandra

- Avro

- YARN

- Spark

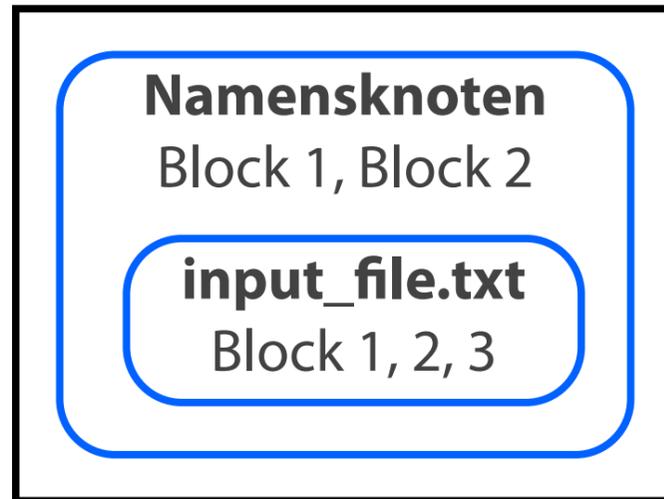
Die Implementierung von MapReduce in Apache Hadoop ist in vielen Bereichen transparent und weist ein stark ausgeprägtes Ecosystem auf.

MapReduce

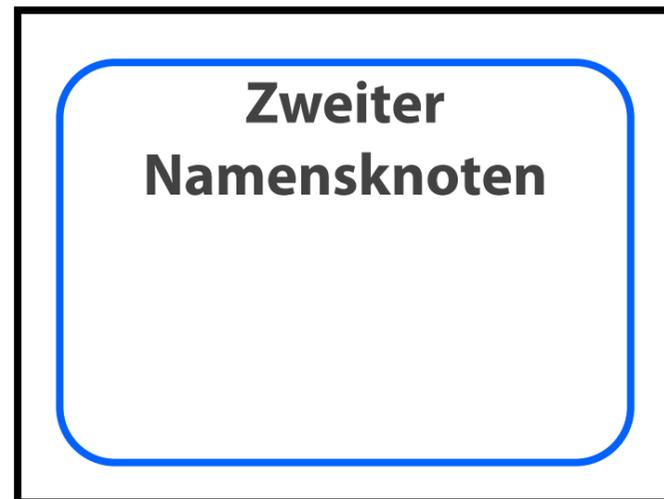
Hadoop Distributed File System (HDFS)

Eine Datei gespeichert im HDFS:

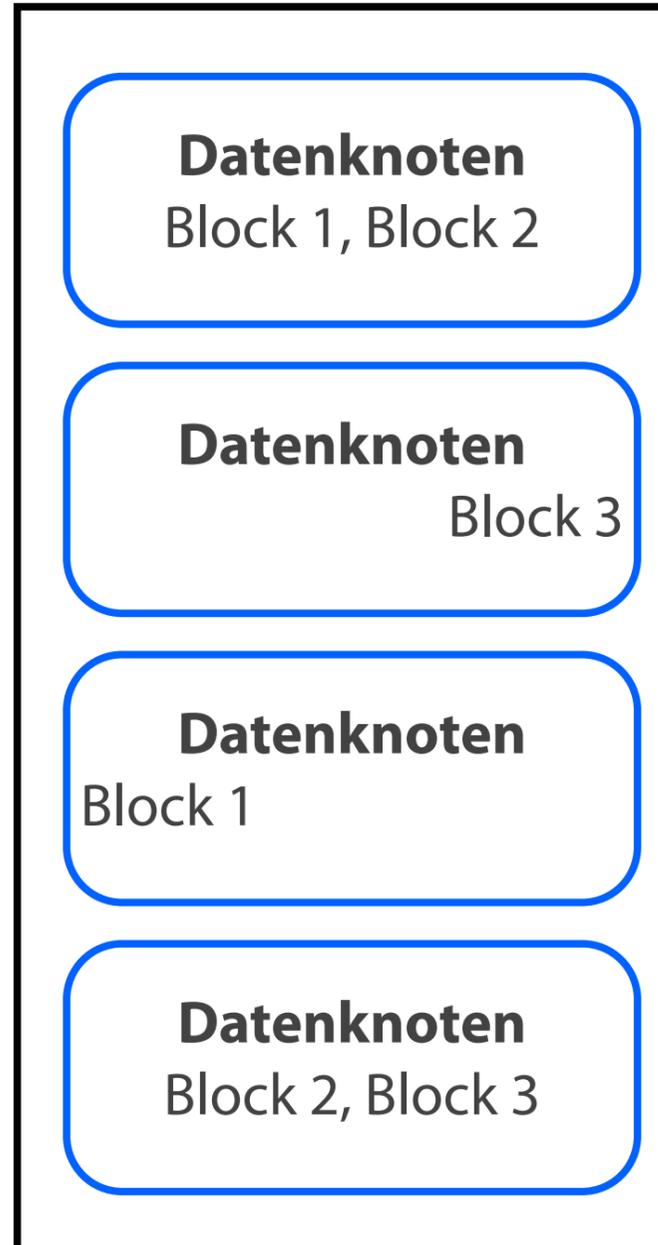
Maschine 1



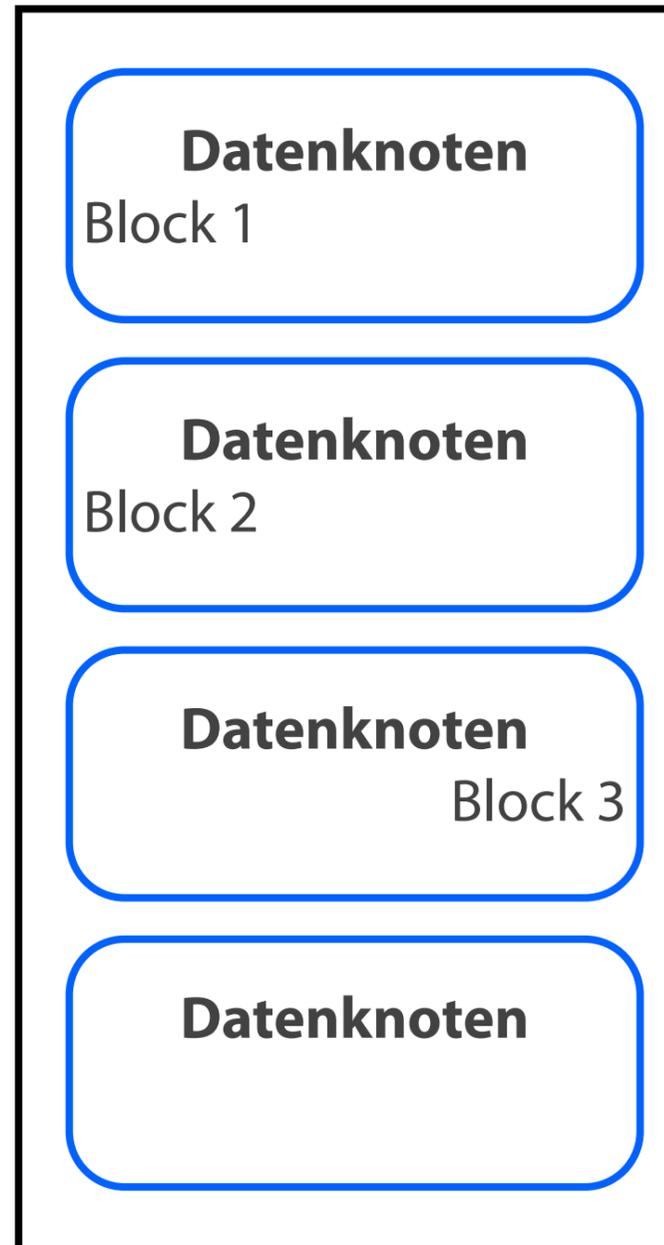
Maschine 2



Maschine 3 -6



Maschine 7-10



Das Hadoop Distributed File System ermöglicht mit Hilfe dreier verschiedener Knotenarten das BigData-gerechte Speichern von Daten und ist Grundlage für den MapReduce-Algorithmus.

MapReduce

Bsp: Qualitätsfehleranalyse

Datensammlung

Maschine	Q4	Q1	Q2	Qual.-fehler
A	1	0	0	1
B	0	1	0	0
A	1	1	0	0
A	1	1	1	1
B	0	0	0	1
A	0	0	0	1
A	0	1	1	0
A	1	0	1	0
C	0	0	1	0
B	1	1	0	1
C	0	1	0	0
A	1	0	0	1
A	0	0	1	1
D	1	0	0	0

Map

Block 1

Qual.-fehler
1
0
1

Key: Q4

Qual.-fehler
0
0
1
0

Key: Q1

Qual.-fehler
1
0

Key: Q2

Block 2

Qual.-fehler
0
1
1
0

Key: Q4

Qual.-fehler
1
0

Key: Q1

Qual.-fehler
0
0
1

Key: Q2

Shuffle & Sort

% Qu.-f.	Anzahl	Key: Q4
66,6	3	
50,0	4	

% Qu.-f.	Anzahl	Key: Q1
25,0	4	
50,0	3	

% Qu.-f.	Anzahl	Key: Q2
50,0	2	
33,3	3	

Reduce

Key	% Qual.-fehler
Q4	57,1
Q1	35,7
Q2	39,9

Die Ausgestaltung eines effizienten MapReduce-Algorithmuses obliegt der Kreativität und dem Wertesystem des Data Scientists.



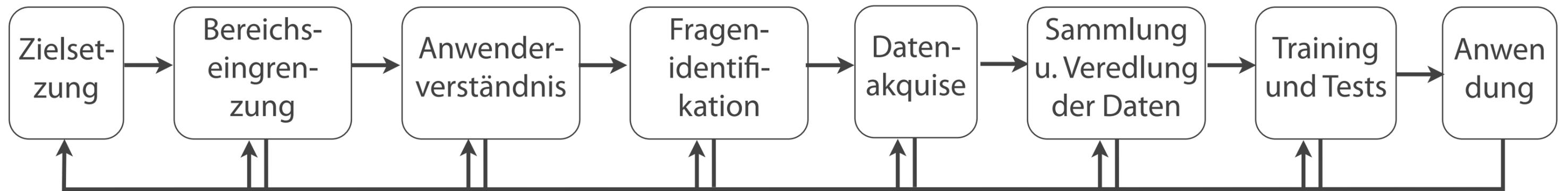
Business Intelligence

Business Analytics

Umgang mit Big Data

Anwendungsvoraussetzungen

Erstellungsprozess für Business Analytics

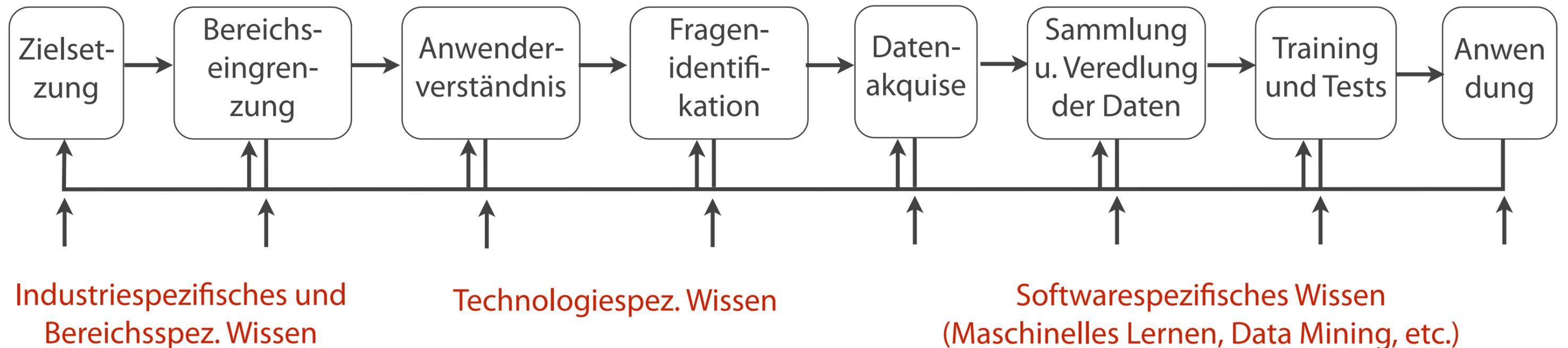


Legende:

Prozess

Nicht erfüllte Voraussetzungen äußern sich in Defiziten entlang des Prozesses zur Erstellung von BA-Systemen.

Voraussetzungen für Business Analytics

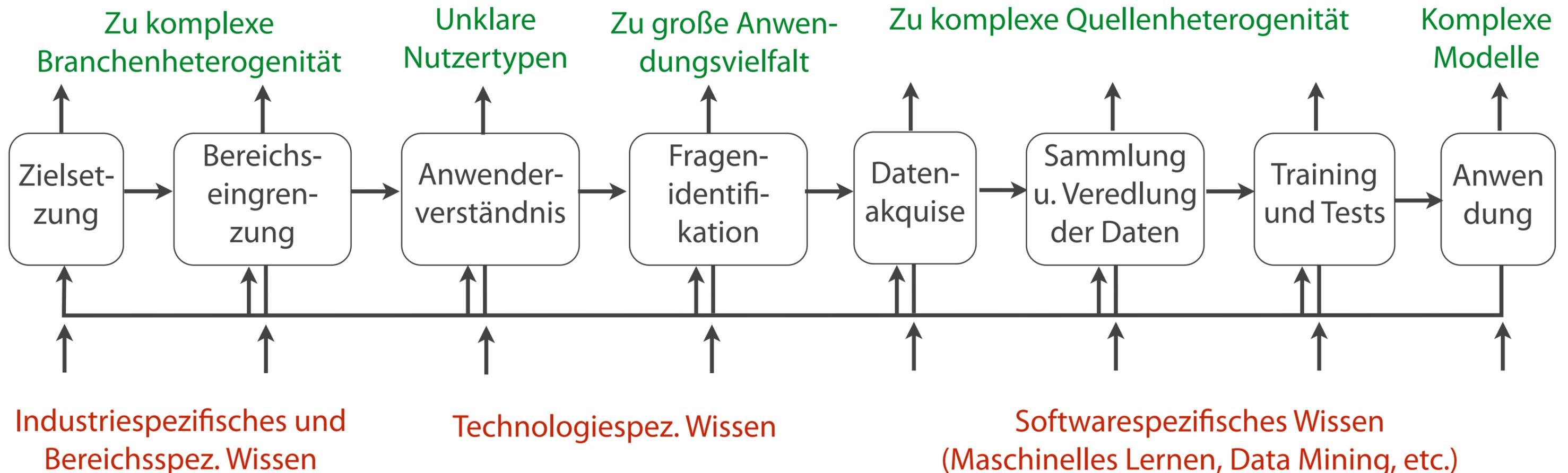


Legende:

Prozess, Voraussetzungen

Nicht erfüllte Voraussetzungen äußern sich in Defiziten entlang des Prozesses zur Erstellung von BA-Systemen.

Herausforderungen von BA

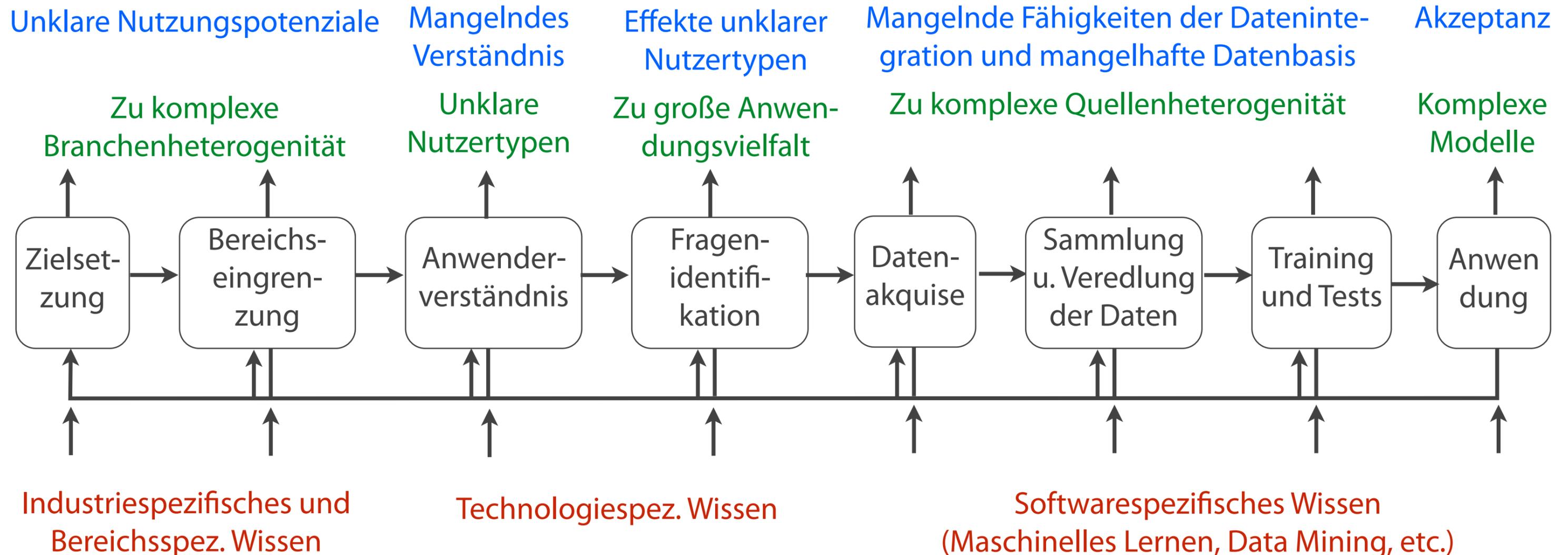


Legende:

Anwender, Entwickler, Prozess, Voraussetzungen

Nicht erfüllte Voraussetzungen äußern sich in Defiziten. Defizite aus frühen Prozessphasen mindern Ergebnisse späterer Prozessphasen zusätzlich.

Herausforderungen von BA



Legende:

Anwender, Entwickler, Prozess, Voraussetzungen

Nicht erfüllte Voraussetzungen äußern sich in Defiziten. Defizite aus frühen Prozessphasen mindern Ergebnisse späterer Prozessphasen zusätzlich.

Literatur

Apache Hadoop 2.7.2 Community: MapReduce Tutorial, <https://hadoop.apache.org/docs/current/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html> 2016

Ashton, K.: That Internet of Things Thing. RFID Journal, 2009

Barbosa, V.C.: An Introduction to Distributed Algorithms, The MIT Press, Cambridge, Massachusetts, London, England, 1996

EMC Education Services: Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, John Wiley & Sons Inc. 2015

Francesca Perino: Tackling Big Data with MATLAB. The MathWorks Inc., 2014

Gronau, N.; Weber, N.; Fohrholz, C.: Wettbewerbsfaktor Analytics - Reifegrad ermitteln, Wirtschaftlichkeitspotenziale entdecken. Gito-Verlag, Potsdam, 2013

Gronau, N.: Analytic Manufacturing. Productivity Management 17. Jg., Heft 5, 2012

Gronau, N.: Analytic Manufacturing. ERP Management 10. Jg., Heft 5, 2014

Hennessy, J.L.; Patterson D.A.: Computer Architecture - A Quantitative Approach. 4. Auflage, 2007

Hurwitz, J.S.; Kaufman, M.; Bowles, A.: Cognitive Computing and Big Data Analytics, John Wiley & Sons Inc., 2015

Kohavi, R.; Rothleder, N.J.; Simoudis, E.: Emerging Trends in Business Analytics. Communications of the ACM, Volume 45, No. 8, S. 45-48, 2002

Nonaka, I.; Takeuchi, H.: The Knowledge-Creating Company - How Japanese Companies Create the Dynamics of Innovation. New York, Oxford University Press., 1995

Scheer, A.-W.: Industrie 4.0 - Wie sehen Produktionsprozesse im Jahr 2020 aus? IMC AG, 2013

Sabherwal, R.; Becerra-Fernandez, I.: Business Intelligence: Practices, Technologies and Management. John Wiley & Sons Inc., 2011

Zum Nachlesen



Grum, M.; Sultanow, E., Friedmann, D.; A. Ullrich, A., Gronau, N.: Tools des Maschinellen Lernens: Marktstudie, Anwendungsbereiche und Lösungen der Künstlichen Intelligenz. Berlin 2021



Gronau, N.; Weber, N.; Fohrholz, C.: Wettbewerbsfaktor Analytics - Reifegrad ermitteln, Wirtschaftlichkeitspotenziale entdecken. Berlin 2013, ISBN: 978-3-95545-051-9

Zum Nachlesen



Kontakt

Dr.-Ing. Marcus Grum

Universität Potsdam
August-Bebel-Str. 89 | 14482 Potsdam
Germany

Tel. +49 331 977 3232
E-Mail mgrum@lswi.de

Gronau, N.:
Industrial Internet of Things – Grundlagen
Berlin 2018, ISBN 978-3955452476 und 978-3955452612

