

IBM BusinessConnect
Vernetzter, intelligenter und informierter denn je



Data Science & Big Data, made in Switzerland

Thilo Stadelmann, ZHAW School of Engineering

Frank Ihringer, Serwise AG

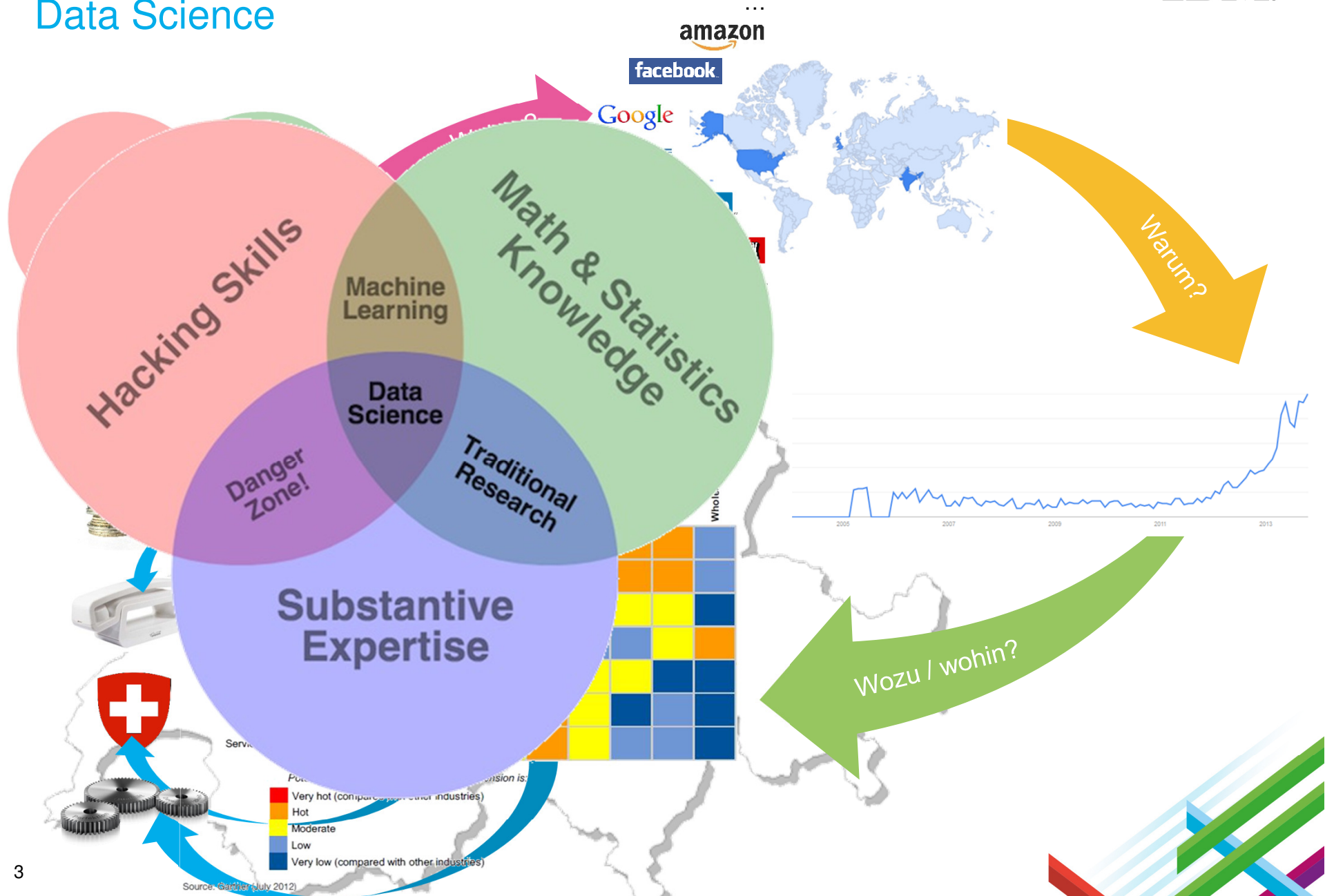


Agenda

- Data Science – made in Switzerland
- Case Study 1: Social Media Mining
- Case Study 2: Text Analytics und Netzwerkvisualisierung
- Case Study 3: Analytisches CRM
- Case Study 4: Clickstream Analyse
- Big Data Roadshow



Data Science



Source: Gartner, July 2012

...made in Switzerland!

Zürcher Hochschule
für Angewandte Wissenschaften

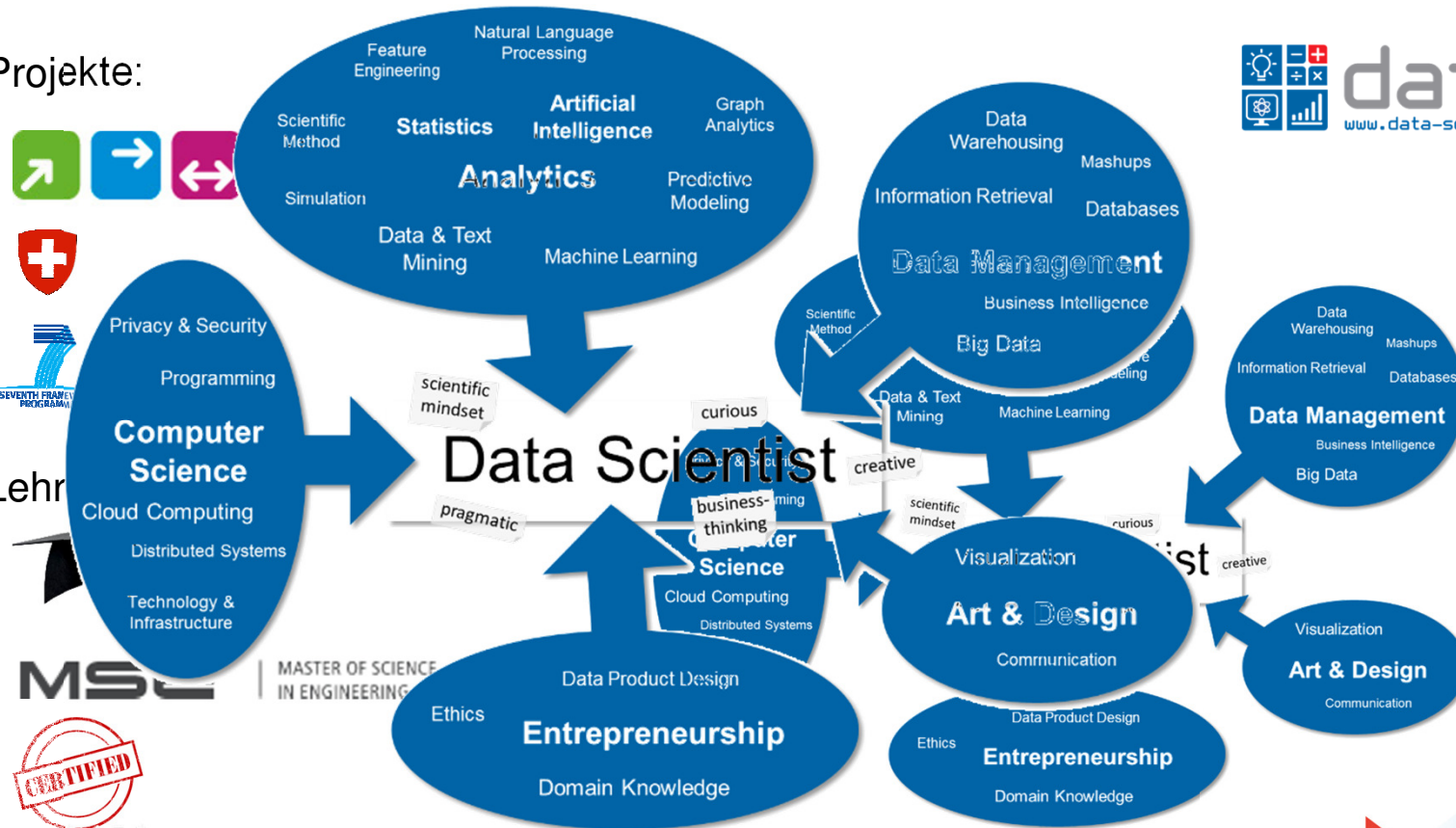


- Datalab: Eines der ersten europäischen Labore für Data Science F&E
- Interdisziplinäres Lab aus Experten der Institute InIT und IDP

• Projekte:



• Lehr



Case Study 1: Social Media Mining



- Aufgabenstellung: «Contextual News for Structured Data»
- Herausforderung: Überbrückung der «Semantic Gap»

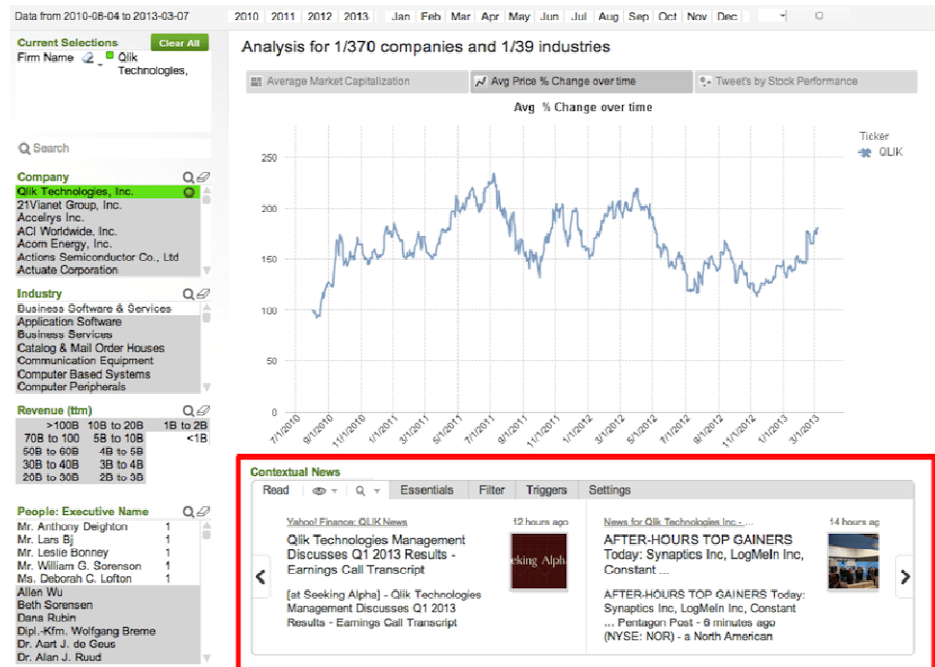


„Ihr habt sie doch nicht alle!“

zufriedenheit

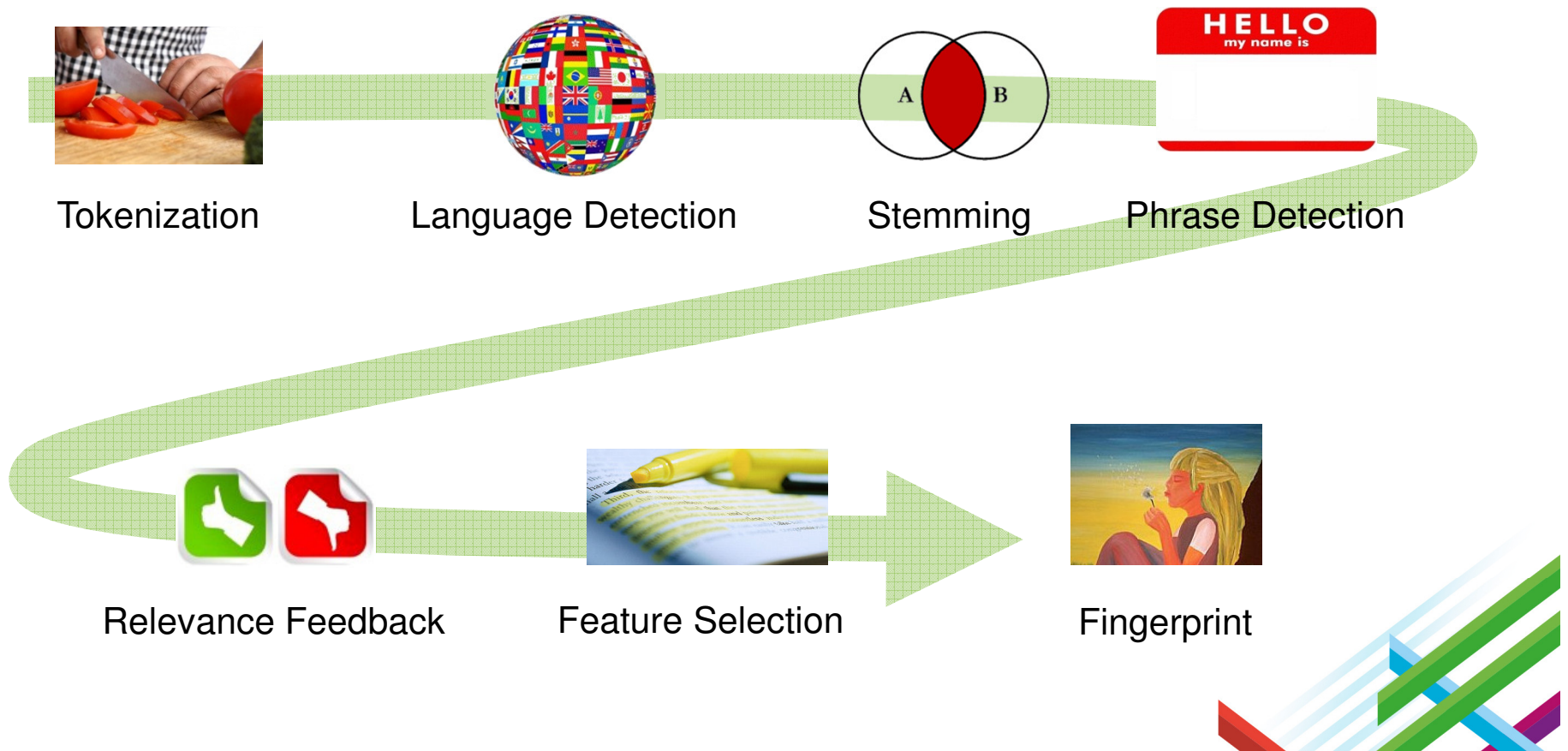
=

low



Case Study 1: Technischer Hintergrund

- Information Retrieval Pipeline:



Case Study 2: Text Analytics und Netzwerkvisualisierung

stiftungschweiz.ch



NonproCons

Neue Wege für Nonprofit-Organisationen

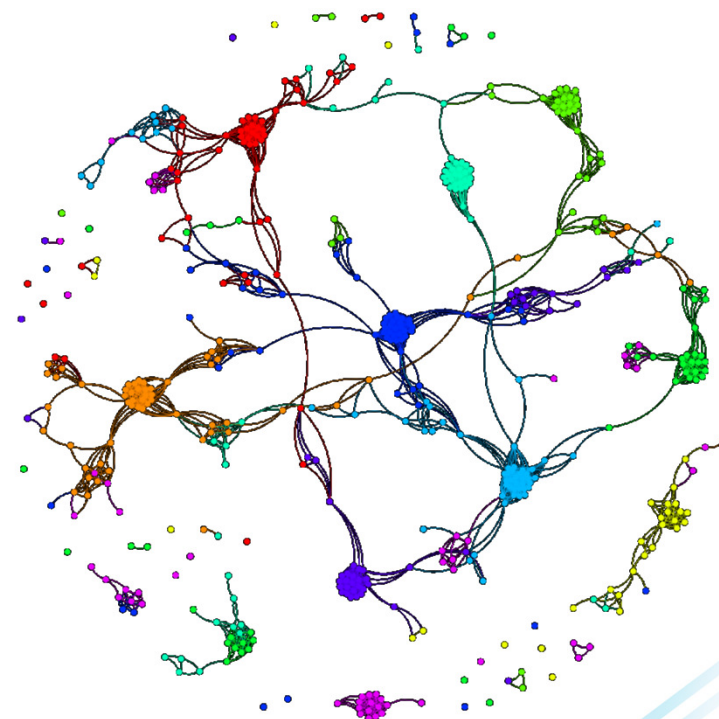
- Ausgangslage: Ca. 7'500 Förderstiftungen in der Schweiz
 - Aufgabe: Proposals & passende Förderer zusammenbringen
 - Herausforderungen:
 - Inhaltliche Ähnlichkeit von Stiftungszwecken quantifizieren
 - Visualisierung der Stiftungslandschaft anhand inhaltlicher Nähe
- *Einfache* Suche anhand Projektskizze nach potentiell bestem Förderer ermöglichen



Case Study 2: Technischer Hintergrund

- Lösungsansätze:
 - Entwicklung mehrsprachiger Retrievalverfahren
 - ➔ BabelNet 2.0: Verwendung von Wikipedia & Wordnet
 - Für Suche extrem kleine Datenbasis (7.5k Stiftungen)
 - ➔ Umsetzung extrem Recall-orientierter Suche
 - Für intuitiv erfassbare Visualisierung sehr viele Daten
 - ➔ Forced Directed Layout auf Ähnlichkeiten aus Term-Dokument-Matrix und Topic Modelling

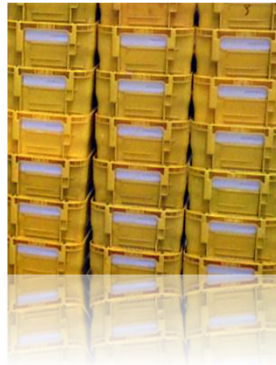
forschung wissenschaftlich basel universität schweiz
 deren forschung unterstützt schweiz künstler
 projekten sozial menschen ausland bereichen
 menschen ausland ähnlicher zielsetzung not
 tier massnahmen forschung erhaltung ausland
 schweiz zürich menschen behindert gemäss
 kinder kindern jugendlichen bern schweiz
 beitragen forschung schweiz ausrichtung beiträg
 menschen unterstützt beiträg einrichtungen kindern
 zuwendungen schweiz ausland gemeinnützigen ausbildung



Case Study 3: Analytisches CRM



- Aufgabe: Welche Kunden per Marketingaktion anschreiben?
- Herausforderungen:
 - Automatisierung der monatl. Mailings

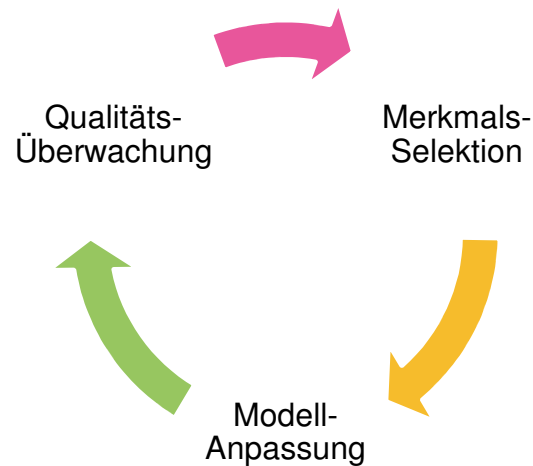


- Nur Response optimieren?

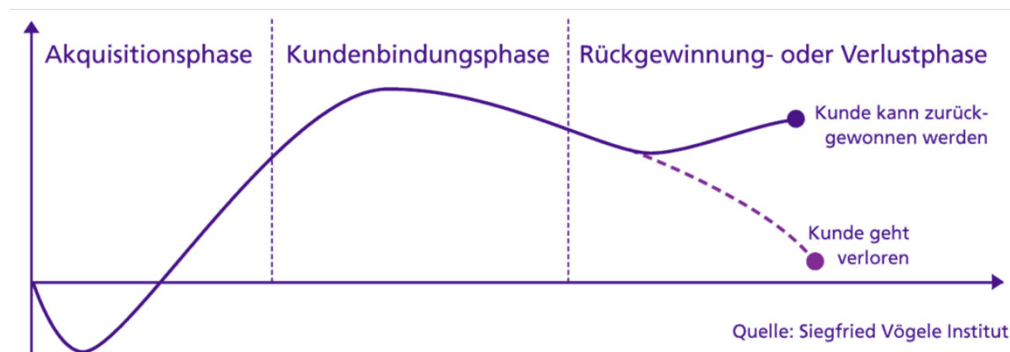


Case Study 3: Technischer Hintergrund

- Lösungsansätze
 - Automatisierung:



- Neue Metrik für Erfolg: Wert eines Kunden auf Lebenszeit betrachtet
 - ➔ Schätzung des Mindset des Kunden (wie entscheidet er wohl in Zukunft?)
 - ➔ Neue Daten notwendig





Case Study 4: Clickstream Analyse

SERWISE
BUSINESS SERVICES

BIG
DATA

- Aufgabe: Welches Wissen schlummert im Serverlog grosser Webportale?

```
[...] [...]
66.249.78.xxx - - [09/Sep/2013:00:00:03 +0200] "GET /product_search?p=12&q=dm+shampoo HTTP/1.1" 200 11659 "-" "Mediapartners-Google" "Mediapartners-Google"
89.244.255.zzz - - [09/Sep/2013:00:00:03 +0200] "GET /WebService/rest/img/id/95809/3 HTTP/1.1" 200 1307 "-" Platform/1.5.1 CFNetwork/609.1.4 Darwin/13.0.0
89.244.255.zzz - - [09/Sep/2013:00:00:03 +0200] "GET /WebService/rest/img/id/95809/3 HTTP/1.1" 200 1764 "-" Platform/1.5.3 CFNetwork/609.1.4 Darwin/13.0.0
89.244.255.zzz - - [09/Sep/2013:00:00:03 +0200] "GET /WebService/rest/img/id/59935/0 HTTP/1.1" 200 1764 "-" Platform/1.5.1 CFNetwork/609.1.4 Darwin/13.0.0
89.244.255.zzz - - [09/Sep/2013:00:00:03 +0200] "GET /WebService/rest/img/id/16551/3 HTTP/1.1" 200 1278 "-" Platform/1.5.1 CFNetwork/609.1.4 Darwin/13.0.0
109.42.3.yyy - - [09/Sep/2013:00:00:03 +0200] "GET /WebService/rest/prod/gsearch/0/4/Estee+Lauder HTTP/1.1" 200 8077 "-" "Apache-HttpClient/UNAVAILABLE (java 1.4)"
[...]
```

- Herausforderung:
 - Datenmenge (Millionen Zeilen pro Tag)
 - Explizit machen von Information (Kontext) → Anreicherung der Logs

- Ideen:
 - Trendspotting á la Google Flu Trends
 - Kundengetriebenes Produktdesign

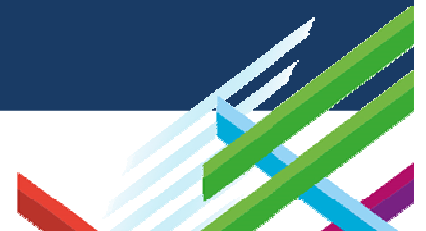


BIG DATA

www.bigdata.ch

Big Data + Business Analytics

Mobiles Testlabor



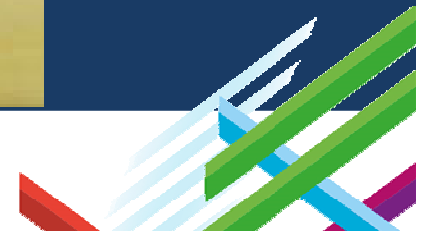
BIG DATA

Feedbacks auf www.bigdata.ch?



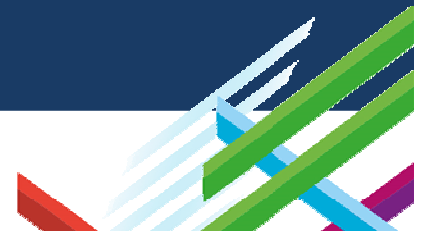
BIG DATA

Feedbacks auf www.bigdata.ch?



BIG DATA

- IBM PureData for Analytics
- IBM Cognos TM1 10.2 OLAP Analytics
- IBM Cognos Express 10.2.1
- IBM Cognos 10 BI
- IBM SPSS Predictive Analytics
- IBM InfoSphere DataStage ETL
- ...
- MS SQL Server
- Oracle DB
- ...



BIG DATA

Business Analytics

Data Warehouse

Data Marts

Predictive Analytics

Business Intelligence

Reporting

FPM Financial Performance

Datenanalyse

Management

Operationelles Reporting

Dashboarding

Close Management

Scorecarding

Budgetierung

Strategische Planung

Operative Planung

Szenarien

Big Data

Web Analytics

Fraud Detection

Risk Management Analytics

Analytical CRM

Social Media Analytics

Sentiment Analytics

Big Data Business Analytics Initiative → Beratungs und Dienstleistungsangebot Serwise Gruppe



BIG DATA

Ein mobiles Test-Center

Ihr Nutzen:

- IBM-Technologie kommt zu Ihnen
- Sie können Ihren Business Case mit allen Software Produkten plus Big Data Server testen.
- Test mit Ihren eigenen Daten, vor Ort
- Einzelner Testtag oder umfangreiches PoC Projekt
- Guter Business Case → 1 Tag kostenlos

Unser Commitment für einen Onsite-Test:

- Die Infrastruktur ist spätestens 30 Minuten nach Ankunft 100 % betriebsbereit.
- Erste Resultate werden nach 4 Stunden demonstriert (z.b. in C10)



BIG DATA Wettbewerbe

Coolste Big-Data-Business-Analytics Studenten-anwendung *)



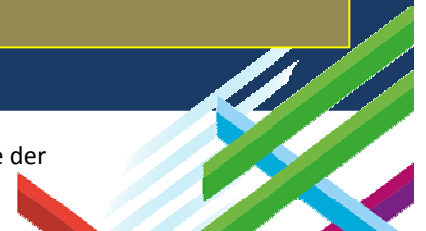
- Die spannendste Big Data Anwendung von Studenten ist gesucht.
- Studenten von Fachhochschulen, Uni, ETH etc. mit Zusammenarbeit der jeweiligen Professoren.

Coolste Big-Data-Business-Analytics Firmen-anwendung *)



- Interessanteste Big Data Anwendung im Enterprise Umfeld ist gesucht.
- Beliebige Unternehmungen und Organisationen
- Fachliche Anforderung respektive ROI Kriterien

Die beschriebene technische Infrastruktur steht von Freitag Abend 18.00 bis Montag morgen 07.00 für Teilnehmer dieser Wettbewerbe am Standort Winterthur zur Verfügung (inklusive Büroräume, Getränke und Kaffeemaschine ...).



Daten aus Netezza in TM1

Administrator: Command Prompt

```
C:\Users\Administrator>"C:\program files\ibm\cognos\tm1_64\bin64\tm1runuti.exe" -process NZ_Cube_Load_Test_1 -server sdata -user admin -pwd apple
```

Administrator: Command Prompt

```
C:\Users\Administrator>"C:\program files\ibm\cognos\tm1_64\bin64\tm1runuti.exe" -process NZ_Cube_Load_Test_2 -server sdata -user admin -pwd apple
```

Administrator: Command Prompt

```
C:\Users\Administrator>"C:\program files\ibm\cognos\tm1_64\bin64\tm1runuti.exe" -process NZ_Cube_Load_Test_3 -server sdata -user admin -pwd apple
```

Administrator: Command Prompt

```
C:\Users\Administrator>"C:\program files\ibm\cognos\tm1_64\bin64\tm1runuti.exe" -process NZ_Cube_Load_Test_4 -server sdata -user admin -pwd apple
```

Administrator: Command Prompt

```
C:\Users\Administrator>"C:\program files\ibm\cognos\tm1_64\bin64\tm1runuti.exe" -process NZ_Cube_Load_Test_5 -server sdata -user admin -pwd apple
```

Administrator: Command Prompt

```
C:\Users\Administrator>"C:\program files\ibm\cognos\tm1_64\bin64\tm1runuti.exe" -process NZ_Cube_Load_Test_6 -server sdata -user admin -pwd apple
```

Administrator: Command Prompt

```
C:\Users\Administrator>"C:\program files\ibm\cognos\tm1_64\bin64\tm1runuti.exe" -process NZ_Cube_Load_Test_7 -server sdata -user admin -pwd apple
```

Administrator: Command Prompt

```
C:\Users\Administrator>"C:\program files\ibm\cognos\tm1_64\bin64\tm1runuti.exe" -process NZ_Cube_Load_Test_8 -server sdata -user admin -pwd apple
```

Resource Monitor

Overview CPU Memory Disk Network

CPU 0% CPU Usage 100% Maximum Freq...

Image	PID	Descri...	Status	Threads	CPU
cmd.exe	3792	Wind...	Running	1	0
cmd.exe	3872	Wind...	Running	1	0
sharedmemoryreader....	2372	share...	Running	3	0
conhost.exe	2580	Conso...	Running	1	0
sharedmemoryreader....	3556	share...	Running	3	0
conhost.exe	3240	Conso...	Running	1	0
tm1_service_launcher....	3340	tm1_s...	Running	3	0
cmd.exe	1972	Wind...	Running	1	0
cmd.exe	2148	Wind...	Running	1	0

Disk 0 KB/sec Disk I/O 0% Highest Active Ti...

Network 21 Kbps Network I/O 0% Network Utilization

Windows Task Manager

Applications Processes Services Performance Networking Users

CPU Usage 0%

CPU Usage History

Memory 6.66 GB

Physical Memory Usage History

Physical Memory (MB)		System	
Total	32767	Handles	21744
Cached	940	Threads	646
Available	25947	Processes	74
Free	25020	Up Time	0:04:06:54
		Commit (GB)	6 / 63

Kernel Memory (MB)

Paged	97
Nonpaged	37

Resource Monitor...

Processes: 74 CPU Usage: 0% Physical Memory: 20%

tm1top: -adminhost localhost -servername sdata -refresh 2

Cmd Id	Username	State	Function
2172	Th:Pseudo	Idle	-
2196	Th:DynamicConfig	Idle	-



Daten aus Netezza in TM1

Administrator: Command Prompt
C:\Users\Administrator>"C:\program files\ibm\cognos\tm1_64\bin64\tm1run.exe" -
process NZ_Cube_Load_Test_1 -server sdata -user admin -pwd apple

Administrator: Command Prompt
C:\Users\Administrator>"C:\program files\ibm\cognos\tm1_64\bin64\tm1run.exe" -
process NZ_Cube_Load_Test_2 -server sdata -user admin -pwd apple

Administrator: Command Prompt
C:\Users\Administrator>"C:\program files\ibm\cognos\tm1_64\bin64\tm1run.exe" -
process NZ_Cube_Load_Test_3 -server sdata -user admin -pwd apple

Administrator: Command Prompt
C:\Users\Administrator>"C:\program files\ibm\cognos\tm1_64\bin64\tm1run.exe" -
process NZ_Cube_Load_Test_4 -server sdata -user admin -pwd apple

Administrator: Command Prompt
C:\Users\Administrator>"C:\program files\ibm\cognos\tm1_64\bin64\tm1run.exe" -
process NZ_Cube_Load_Test_5 -server sdata -user admin -pwd apple

Administrator: Command Prompt
C:\Users\Administrator>"C:\program files\ibm\cognos\tm1_64\bin64\tm1run.exe" -
process NZ_Cube_Load_Test_6 -server sdata -user admin -pwd apple

Administrator: Command Prompt
C:\Users\Administrator>"C:\program files\ibm\cognos\tm1_64\bin64\tm1run.exe" -
process NZ_Cube_Load_Test_7 -server sdata -user admin -pwd apple

Administrator: Command Prompt
C:\Users\Administrator>"C:\program files\ibm\cognos\tm1_64\bin64\tm1run.exe" -
process NZ_Cube_Load_Test_8 -server sdata -user admin -pwd apple

Windows Task Manager
Applications | Processes | Services | Performance | Networking | Users
CPU Usage | CPU Usage History

Resource Monitor
File Monitor Help
Overview | CPU | Memory | Disk | Network
CPU 0% CPU Usage
Disk 0 KB/sec Disk I/O
Network 21 Kbps Network I/O

CubeViewer: SData->NZ_Testcube_Netezza_Load->Standard [Privat]
Datei Bearbeiten Ansicht Optionen Hilfe
Standard [Base]
TOTAL PRODUCT TOTAL CUSTOMER

NZ_Day	NZ_Measure						
	Order Quantity	Extend Amount	Product Standard Cost	Total Product Cost	Sales Amount	Tax Amount	Freight
+ 2006	48'719	119'131'114.29	70'806'160.19	70'806'160.19	119'131'114.29	9'530'489.55	2'978'279.80
+ 2007	443'310	177'810'222.96	103'934'323.44	103'934'323.44	177'810'222.96	14'224'818.33	4'445'275.08
+ 2008	558'397	170'248'561.84	99'857'732.55	99'857'732.55	170'248'561.84	13'619'884.95	4'256'239.31
+ 2026	1'656'037	507'958'834.65	296'659'808.36	296'659'808.36	507'958'834.65	40'636'706.77	12'699'045.16
+ 2066	68'065	166'030'244.94	98'753'003.45	98'753'003.45	166'030'244.94	13'282'420.14	4'150'758.84
+ 2067	619'372	249'827'137.61	145'975'562.65	145'975'562.65	249'827'137.61	19'986'171.66	6'245'705.59
+ 2068	807'442	245'645'564.81	143'780'359.59	143'780'359.59	245'645'564.81	19'651'645.18	6'141'175.56
+ 2086	1'366'971	411'915'079.26	241'222'676.95	241'222'676.95	411'915'079.26	32'953'206.34	10'297'938.70
+ 2107	23'637	76'111'419.34	45'559'925.47	45'559'925.47	76'111'419.34	6'088'913.66	1'902'786.65
+ 2108	61'710	151'539'035.24	90'133'907.81	90'133'907.81	151'539'035.24	12'123'123.32	3'788'478.35
+ 2109	557'750	224'672'098.47	131'372'894.71	131'372'894.71	224'672'098.47	17'973'768.48	5'616'826.97
+ 2110	725'685	219'231'103.25	128'531'104.32	128'531'104.32	219'231'103.25	17'538'488.26	5'480'810.43
+ 2126	63'173	154'415'495.56	91'806'602.87	91'806'602.87	154'415'495.56	12'353'240.17	3'860'389.91
+ 2127	574'386	231'299'917.45	135'213'976.23	135'213'976.23	231'299'917.45	18'503'994.02	5'782'523.11
+ 2128	748'286	226'894'097.58	132'988'314.07	132'988'314.07	226'894'097.58	18'151'527.81	5'672'386.23
+ 2146	1'132'313	348'395'774.75	203'336'998.28	203'336'998.28	348'395'774.75	27'871'661.98	8'709'945.11

23'637

8 Mio Factrows in 2.5 Min
56 Mio Intersektionen geschrieben → 375'000 Intersektionen pro Sekunde
(3 Jahre alter Server, nur lokale Disks, kein SAN, kein Fibre Channel zu Storage)
→ Potentiell: Faktor 4 bis 8 schneller ...

DE 22:35 12.09.2013



BIG DATA

BIG DATA

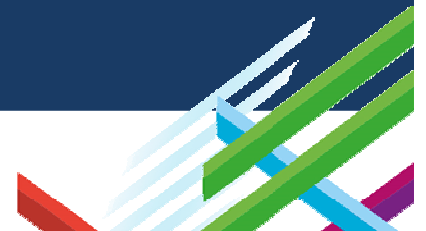
Business Analytics Wettbewerb

Machen Sie mit bei unserem **BIG DATA Wettbewerb!**

Beantworten Sie folgende Fragen.

Unsere Mitarbeitenden mit den «ASK ME» Buttons helfen Ihnen sehr gerne bei der Beantwortung der Wettbewerbsfragen.

Viel Spass beim Knobeln und Diskutieren.



BIG DATA

Kontakt

Serwise AG

Seidenstrasse 2, 8400 Winterthur

Phone +41 (0) 52 212 68 34, Fax +41 (0) 52 212 68 96

Frank Ihringer (Projektleiter), frank.ihringer@serwise.com

Philipp Altmann (Technischer Kontakt), philipp.altmann@serwise.com

Matthias Zimmermann (Organisation), matthias.zimmermann@serwise.com

NEHMEN SIE KONTAKT MIT UNS AUF, WIR FREUEN UNS

