



TERADATA®

Key Trends in Big Data and Analytics

Martin Willcox, Director Big Data Centre of Excellence (Teradata International)

October 2015

© 2015 Teradata



Agenda

- Motivating examples from an “old” industry
- From transactions and events to interactions and observations
- Key Trends, a.k.a.: Tomorrow’s Big Data & Analytics
- Summary and conclusions

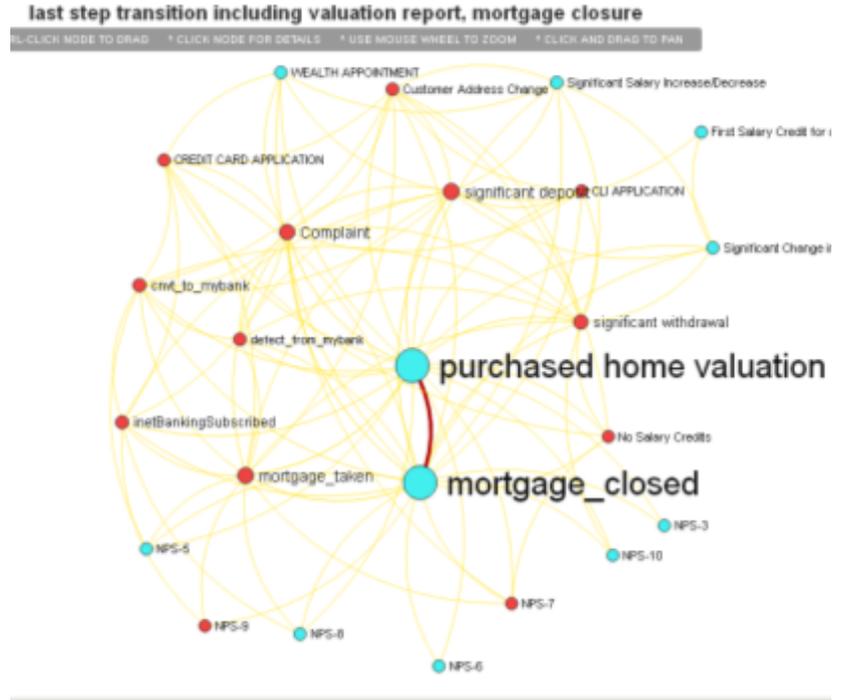
The background of the slide is a complex digital visualization of financial market data. It features multiple overlapping candlestick charts in shades of blue and white. Interspersed among the charts are various numerical values, including stock prices like 3225.89, 2956.21, 6725.89, 8956.21, 2540.44, 1235.41, 1125.25, and 3256.41. There are also percentage changes such as +11.0, 0.00, .02%, and -0.02%, along with other indicators like 55.02, 10, 01, and 02:11. The overall aesthetic is high-tech and data-driven, with a grid-like pattern in the background.

Motivating examples from an “old” industry

Preventing mortgage churn

(transaction data, text analytics, fuzzy matching and graph analytics)

How can we identify customers who are about to re-finance their homes and take a mortgage from another Bank?



Managing customer complaints

(call center agent notes, text analytics and fuzzy matching)



12.05.2000 (10:02) XXX (XXX) op-qld telephone bankthis additional
this complaint loan account - XXX

taking his complaint to the banking ombudsman after
this call and was considering taking all his business away from us

```
THORISATION FOR A CHALLENGED TRANSACTION BY ONE - 1607C  
STOMERS.XXXXXX ADV WE SHOULD HAVE WRITTEN TO ADVIS ed to coll  
STOMER WAS CHALLENGING TRANSACTION AS PREVIOUSLY cor  
HAVE DONE. XXXXX ADV HAS THIS SLIP CUSTOMER SIGN  
THORISE THE TRANSACTION. XXXXX WANTS THE $750 PLU  
VERSED OR IS WRITING TO THEOMBUDSMAN.  
LTS TAKEN BY XXXXX  
.01.2005 (12:58) XXXXX (08 8424 8552) SLCRU INBOX  
EA RESP changed from - CRSLIN - SLCRU INBOX  
- CRULIN - ULCRU INBOX  
T changed from - CRSLIN - SLCRU INBOX  
- CRULIN - ULCRU INBOX  
VESTIGATE AS CUSTOMER IS DISPUTING THE TRANSACTIO
```

has changed the phone on a account overdraft by \$210 dollars to a
\$275,000 account loan that is obscene That is taking your rules to a very poor
customer service level I will be looking to move accounts to another Bank as
soon as minimum period is over if this is how you continue to treat customers
enquiryType : Home Loans outcomeSought : I would like my \$40
dollars back userTitle : Mr userTitleOther : userFirstName :

How can we identify customers whose concerns are on the verge of a dangerous escalation?

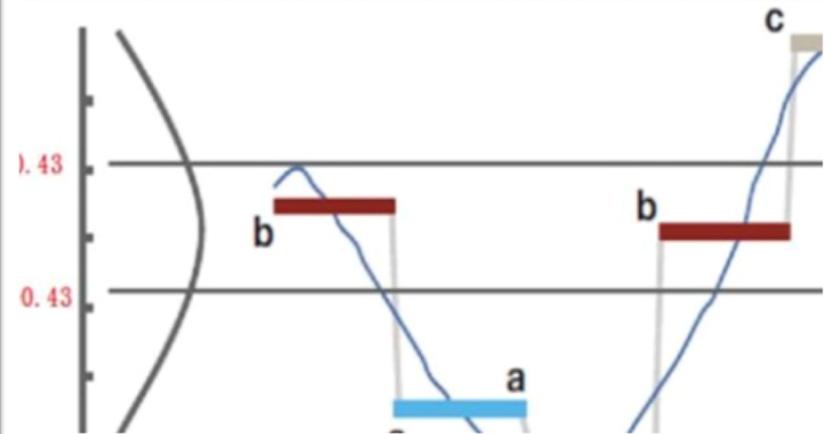
Predicting default – whilst minimising false positives

(account balance movement and time-series Analytics (SAX))



Creating alerts based on large balance movements generates 000s of false alarms...

...sophisticated time-series Analytics and pattern-matching reduces alarm output by 90% whilst increasing predictive accuracy by 3x

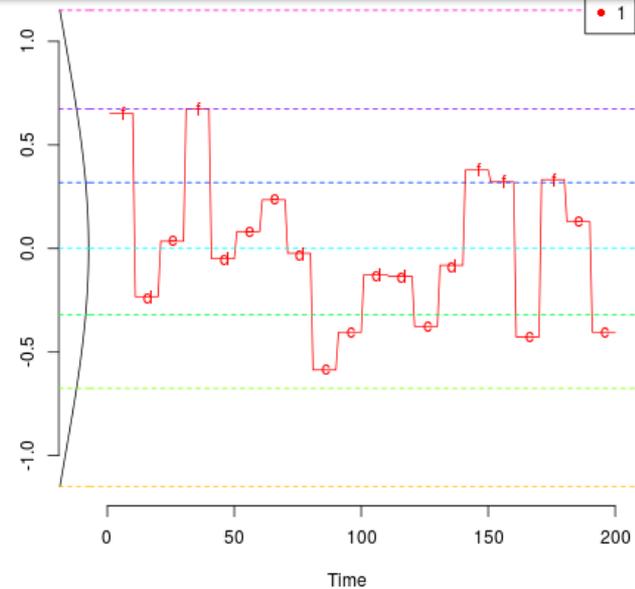


What do all of these examples have in common?

"New" data

```
THORISATION FOR A CHALLENGED TRANSACTION BY ONE
STOMERS.XXXXX ADV WE SHOULD HAVE WRITTEN TO ADVIS
STOMER WAS CHALLENGING TRANSACTION AS PREVIOUSLY
HAVE DONE. XXXXX ADV HAS THIS SLIP CUSTOMER SIGN
THORISE THE TRANSACTION. XXXXX WANTS THE $750 PLU
VERSED OR IS WRITING TO THEOMBUDSMAN.
LTS TAKEN BY XXXXX
.01.2005 (12:58) XXXXX (08 8424 8552) SLCRU INBOX
EA RESP changed from - CRSLIN - SLCRU INBOX
- CRULIN - ULCRU INBOX
T changed from - CRSLIN - SLCRU INBOX
- CRULIN - ULCRU INBOX
VESTIGATE AS CUSTOMER IS DISPUTING THE TRANSACTIO
OM HIS ACCOUNT.
.01.2005 (12:58) XXXXX (08 8424 8552) SLCRU INBOX
```

"New" Analytics



The Three New Waves of the New Big Data

A short history of Big Data

(1)

Analysis of web / clickstream data enables Google, Amazon, eBay and others to achieve “mass customisation”.

People interacting
with things

(2)

Clickstream / social / mobile interaction data enables Amazon, LinkedIn, Netflix, etc. to go social (“people who like what you like also like...”)

People interacting
with people

(3)

Increasing instrumentation is now leading to the emergence and optimisation of “the Internet of Things”.

Things interacting
with things

Riding the three “new waves” of Big Data

From transactions and events - to interactions and observations

**Simple computing devices are now so inexpensive
that increasingly everything is instrumented**

**Instead of capturing transactions and events in the Data Warehouse
and inferring behaviour (of customers and systems), we can
increasingly measure it directly**

**Organisations making the “transactions, to interactions”
journey need to address five key challenges...**

The “Big 5”

The challenges of making the “transactions, to interactions” journey

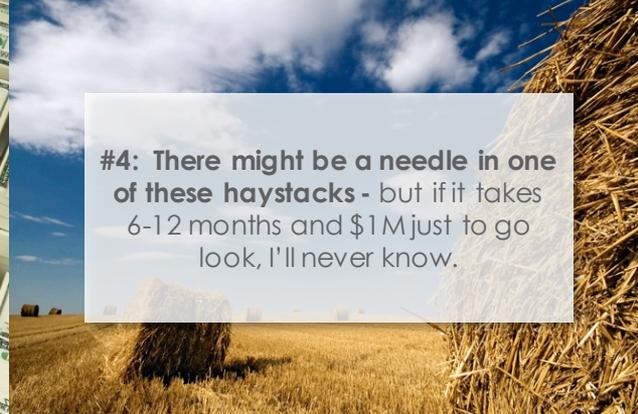


#1: The requirement to manage multi-structured data and data whose structure changes continuously means that there is no single Information Management strategy that works equally well across the entire Big Data space.

#2: Understanding Interactions requires path / graph / time-series Analytics in addition to traditional “set-based” Analytics, so that there isn't a single parallel processing framework or technology that works equally well across the entire Big Data space.



#3: The economic challenge of capturing, storing, managing and exploiting Big Data sets that may be large; getting larger quickly; noisy; of (as yet) unproven value; and infrequently accessed.

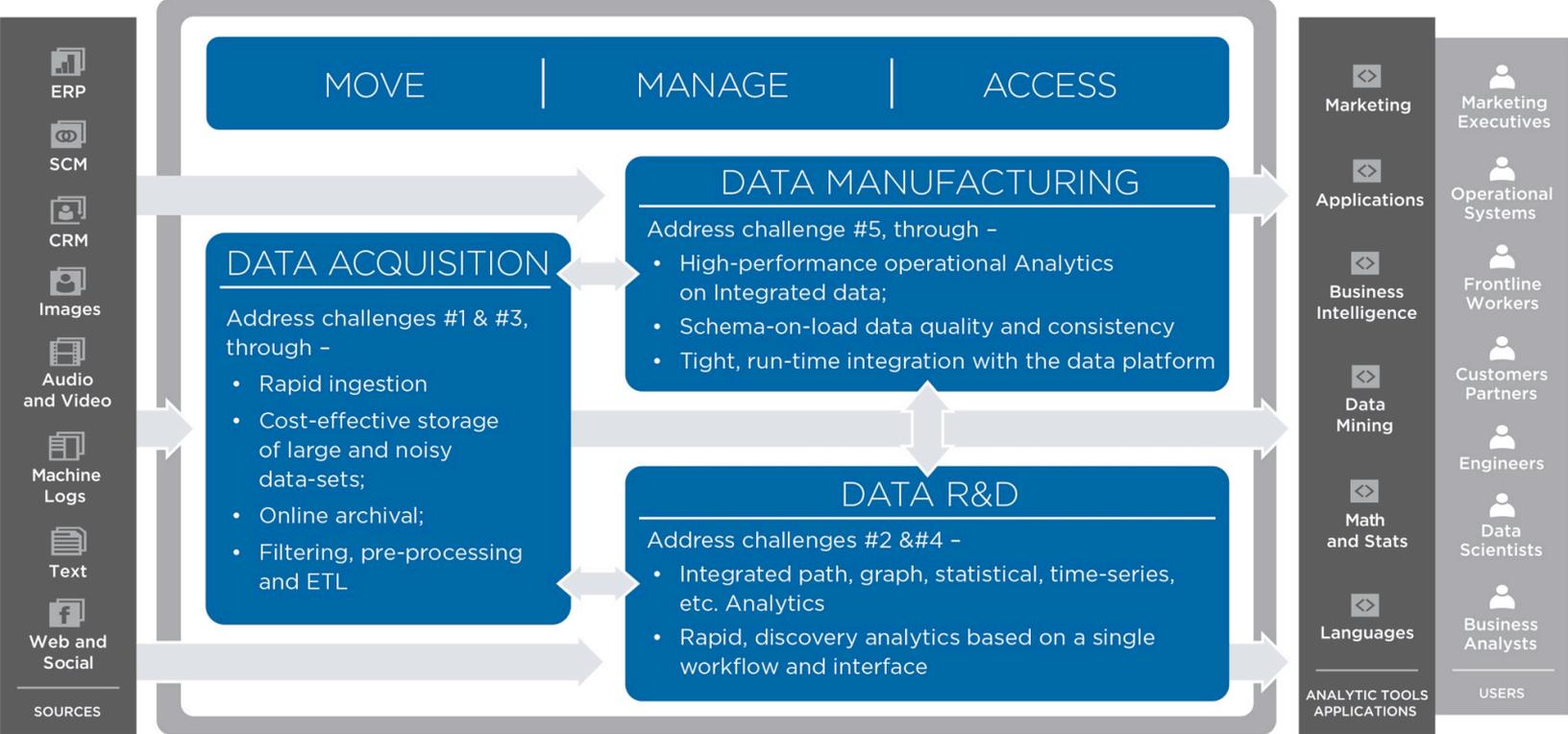


#4: There might be a needle in one of these haystacks - but if it takes 6-12 months and \$1M just to go look, I'll never know.

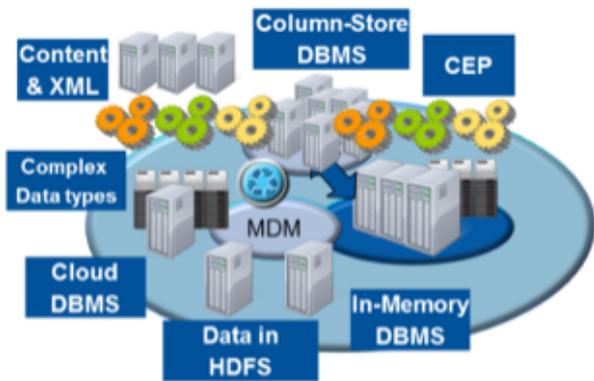


#5: Getting past “so what” to drive real business value (because old business process + expensive new technology = expensive, old business process)

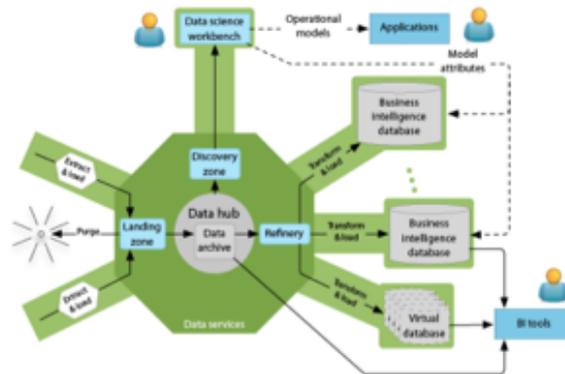
Five challenges, three key capabilities



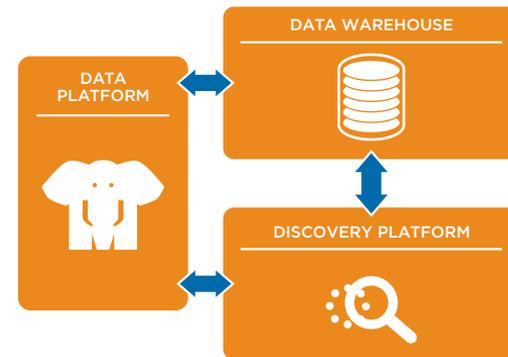
– Gartner –
Logical Data Warehouse



– Forrester –
Enterprise Data Hub



– Teradata –
Unified Data Architecture



*We can't all agree what to call the new Analytic Architecture, but we all agree –
Big Data Are Plural*

Multiple Information
Management strategies -
and multiple technologies

Ecosystem, rather than
purely platform focus

Integration becomes a
critical concern

TERADATA

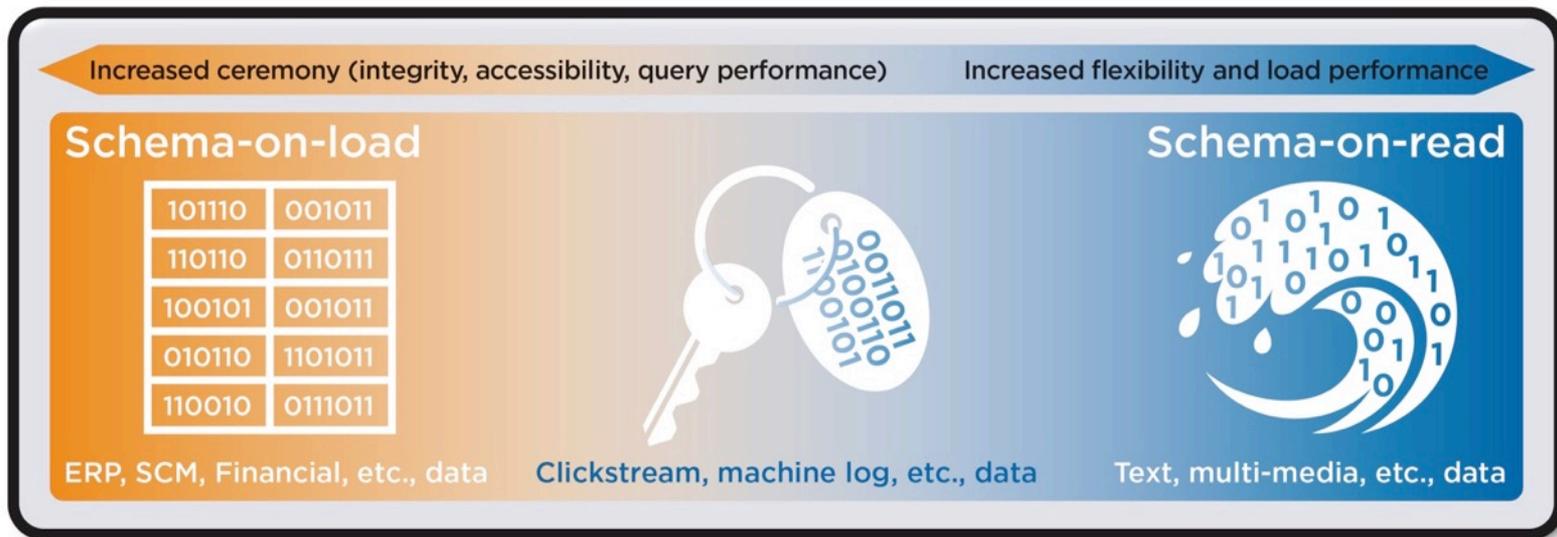


Tomorrow's Big Data & Analytics*

**It is difficult to make predictions, especially about the future*

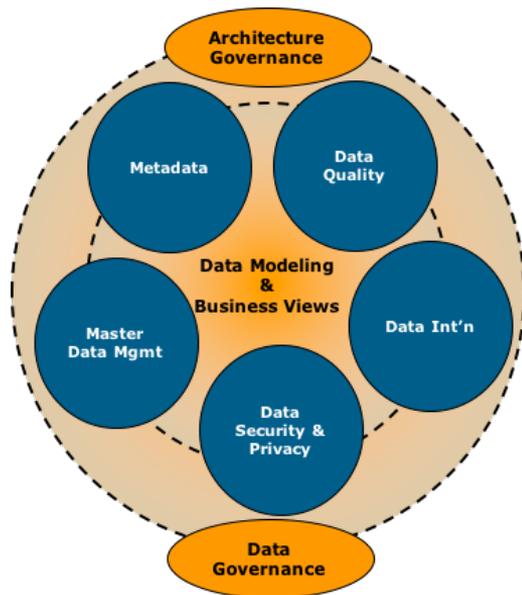


#1 The end of the relational assumption does not mean the end of the relational database / data warehouse



Explicit, or implicit, there is always, always, always (at least one) schema;
“pay me now, or pay me later (and over and over)”

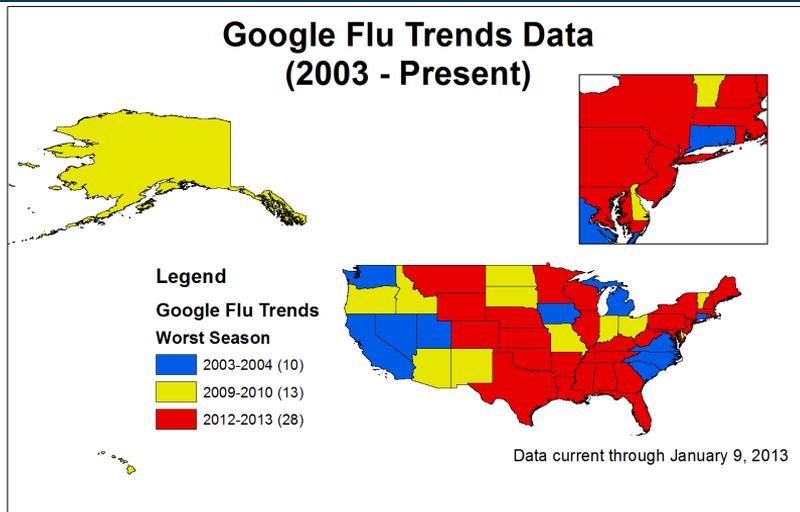
#2 Information Management and Governance models will evolve



#3 The New Analytics will extend, not replace, traditional Analytics

A tale of two "Big Analytics" projects

Either / Or



Ignored existing data-sets and models; consistently over-estimates number of cases and predicts winter, rather than flu.

Both And

The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales

Lynn Wu
The Wharton School
3730 Walnut St. JMH31 561
Philadelphia, PA 19104
wlynn@wharton.upenn.edu

Erik Brynjolfsson
MIT Sloan School of Management
50 Memorial Drive, E53-313
Cambridge, MA 02142
erikb@mit.edu

First draft: March, 2009, This Draft: August, 2013

[Comments Welcome](#)

Abstract

We demonstrate how data from search engines such as Google provide an accurate but simple way to predict future business activities. Applying our methodology to predict housing market trends, we find that a housing search index is strongly predictive of future housing market sales and prices. For state-level predictions in the US, the use of search data produces out-of-sample predictions with a smaller mean absolute error than the baseline model that uses conventional data but lacks search data. Furthermore, we find that our simple model of using search frequencies beat the predictions made by experts from the National Association of Realtors by 23.6% for future US home sales. We also demonstrate how these data can be used in other markets, such as home appliance sales. In the near future, this type of "nanoeconomic" data can transform prediction in numerous markets, and thus business and consumer decision-making.

Keywords: Online Search, Prediction, Housing Prices, Real Estate, Google Trends

[Acknowledgements](#): We thank Karl Case, Avi Goldfarb, Andrea Meyer, Dana Meyer, Shichun Reichman, Lu Han, and Hal Varian as well as seminar participants at the NBER, MIT, the Workshop on Information Systems and Economics, and the International Conference on Information Systems for valuable comments on this research. The MIT Center for Digital Business provided generous funding.

Extended an existing model based on transactional data with new, Google search based features; significant uplift.

#4 Bigger roles for discovery-oriented Analytics and data-discovery will mean embracing “bi-modal IT”

Deeply Different, Both Essential

	Mode 1		Mode 2	
	Reliability	Goal	Agility	
	Price for performance	Value	Revenue, brand, customer experience	
Think Marathon Runner	Waterfall, V-model, high-ceremony IID	Approach	Agile, Kanban, low ceremony IID	Think Sprinter
	Plan driven, approval based	Governance	Empirical, continuous, process based	
	Enterprise suppliers, long-term deals	Sourcing	Small, new vendors, short-term deals	
	Good at conventional process, projects	Talent	Good at new and uncertain projects	
	IT-centric, removed from customer	Culture	Business-centric, close to customer	
	Long (months)	Cycle times	Short (days, weeks)	

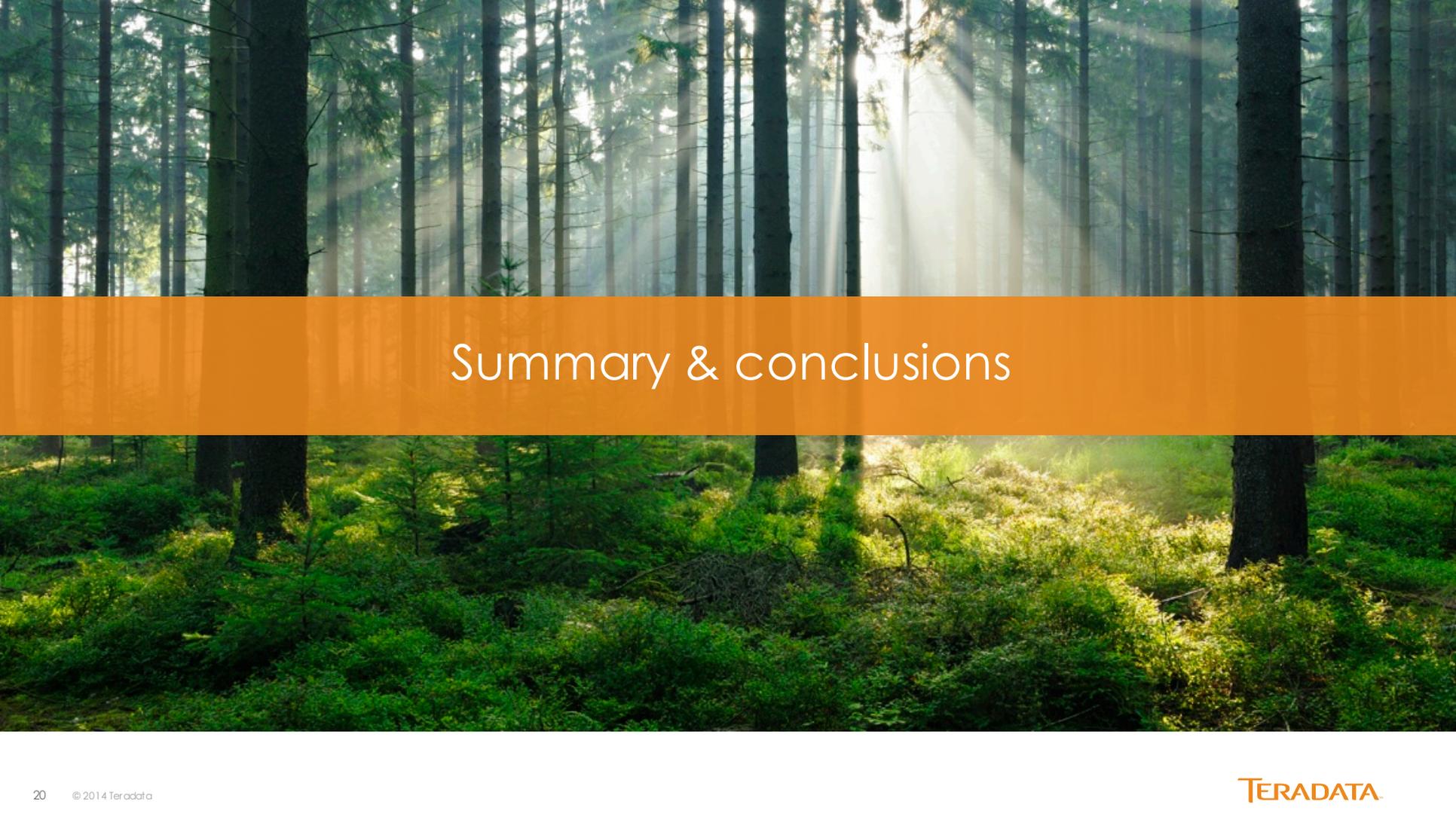
Copyright Gartner Group (2015)

#5 It's about the operationalisation, stupid

Old business process + Expensive new technology = Expensive old business process



The objective is not merely to gain insight – the objective is to change the way we do business.



Summary & conclusions



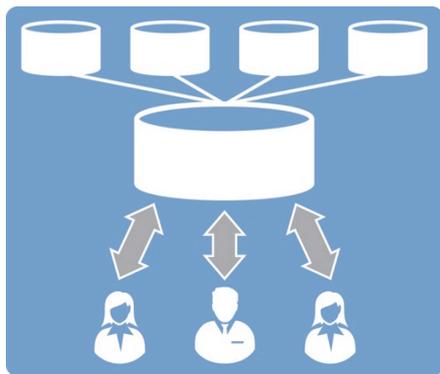
Digitization, in short, is not a great equalizer that drives all companies toward similar processes and outcomes. Instead, it's driving the leaders and laggards further apart.

Andrew McAfee & Erik Brynjolfsson

The “Logical Data Warehouse” is the industry’s adaptation to “Big Data” – plan to deploy

How will you deploy? How many / which platforms will you need? How will you integrate them? Which data need to be tightly / loosely coupled?

The Enterprise Data Warehouse Era

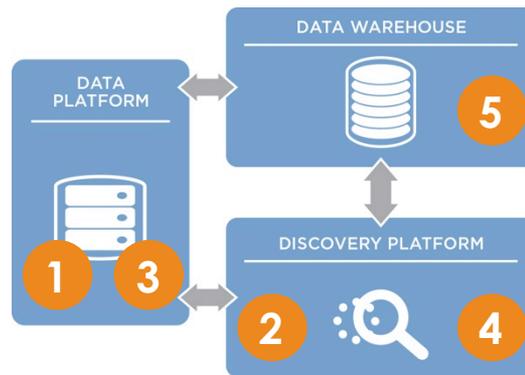


“Give me integrated, high quality data.”



The Logical Data Warehouse (a.k.a.: Unified Data Architecture) Era

- 1 Multi-structured data
- 2 Interaction / observation Analytics
- 3 Flat / falling IT budgets, exploding data volumes
- 4 Agile Exploration & Discovery
- 5 Operationalisation



“Centralise and integrate the data that are widely re-used and shared, but integrate all of the analytics.”

TERADATA®

