

What is ... Big Data?

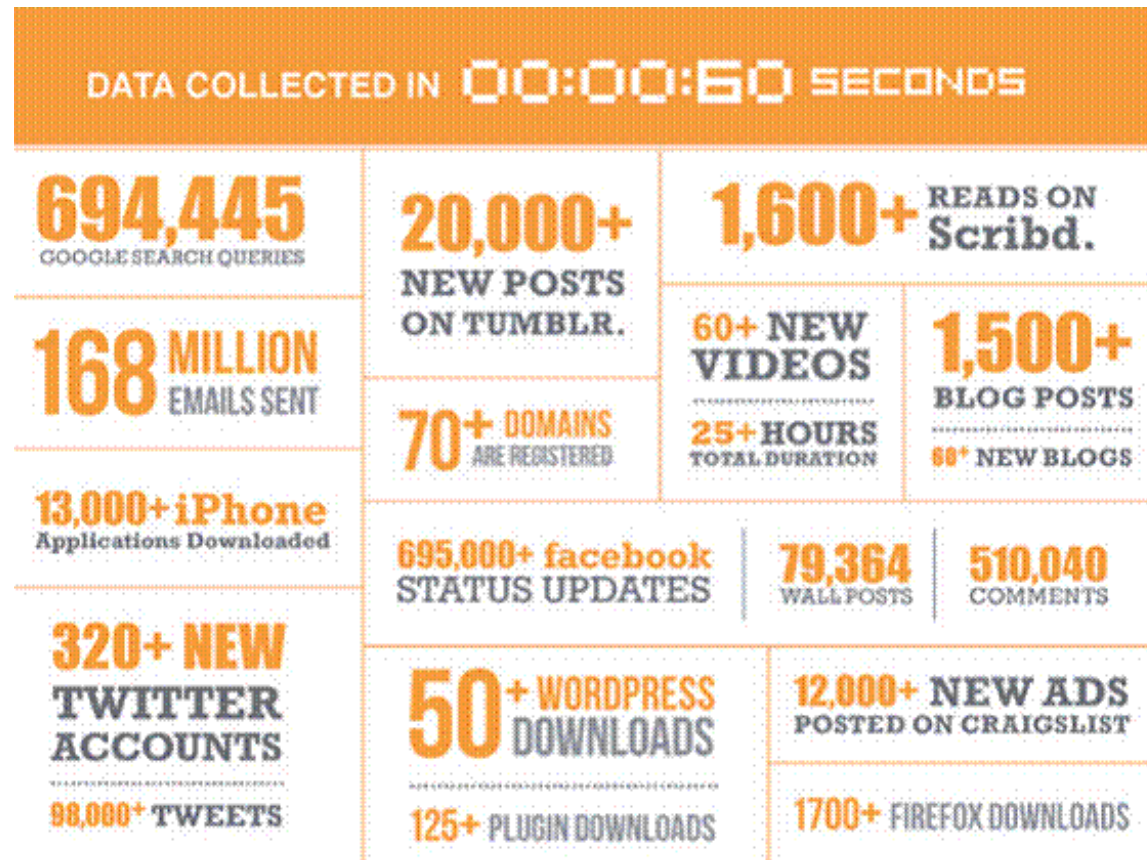
Vincent Jeansoulin
Analytics Platform Specialist

Vincent.jeansoulin@ch.ibm.com

 @vjeansoulin  Vincentjeansoulin



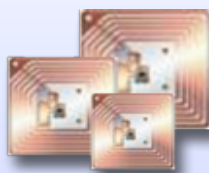
Big Data ... is BIG?



-> Not Only

Big Data ... is FAST?

Responding to the
increasing **Velocity**



30 Billion
RFID sensors
and counting

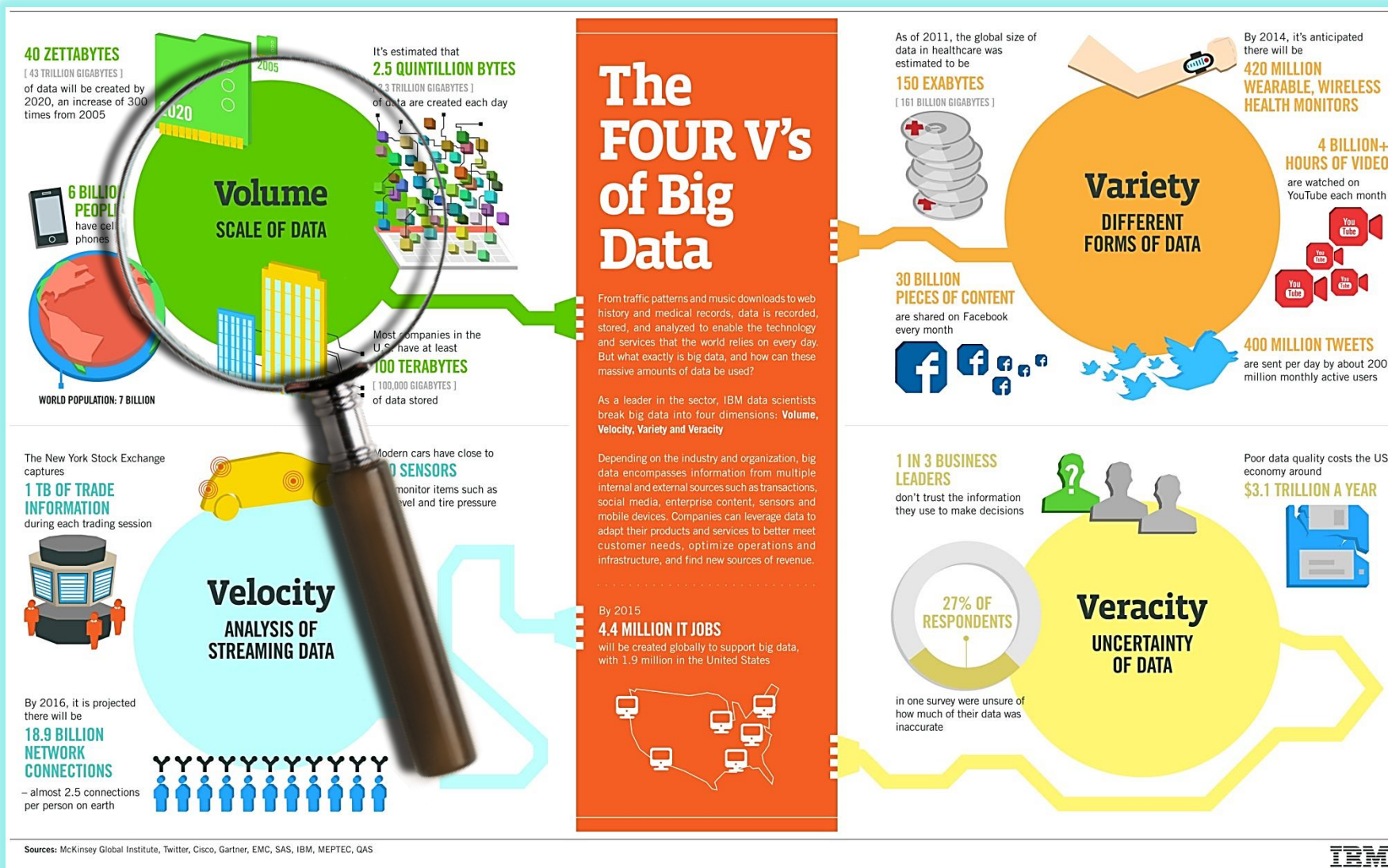
Big Data ... is everything? ... Is everything Big Data?

Collectively analyzing
the broadening **Variety**



80% of the
worlds data is
unstructured

The Four V's



Big Data is ...

- Not about having the data, it is about doing something with data to generate value!
- Driven and fuelled by fundamental changes such as eCommerce, the Internet, data generating gadgets, sensors, ...

Big Data is ... Already part of our life



Big Data ... can help to transform business models



Vestas optimizes capital investments based on 2.5 Petabytes of information.

- Model the weather to optimize placement of turbines, maximizing power generation and longevity.
- Reduce time required to identify placement of turbine from weeks to hours.
- Incorporate 2.5 PB of structured and semi-structured information flows. Data volume expected to grow to 6 PB.



Vestas

Big Data ... can help to transform our daily life



KTH Swedish Royal Institute of Technology Reducing Traffic Congestion

Capabilities Utilized:

Stream Computing

- Deployed real-time Smarter Traffic system to predict and improve traffic flow.
- Analyzes streaming real-time data gathered from cameras at entry/exit to city, GPS data from taxis and trucks, and weather information.
- Predicts best time and method to travel such as when to leave to catch a flight at the airport

Significant benefits:

- Enables ability to analyze and predict traffic faster and more accurately than ever before
- Provides new insight into mechanisms that affect a complex traffic system
- Smarter, more efficient, and more environmentally friendly traffic



Big Data ... can help to create a completely new class of analytical applications



TerraEchos uses streaming data technology to support covert intelligence and surveillance sensor systems

Need

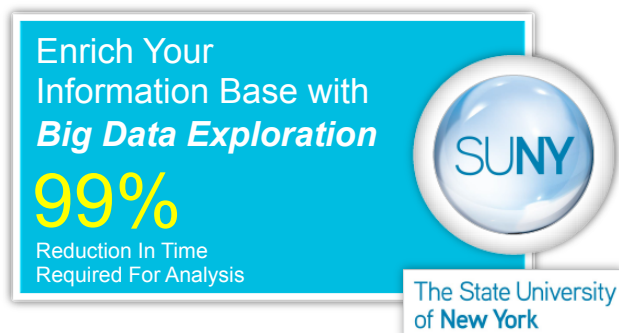
- Deployed security surveillance system to detect, classify, locate, and track potential threats at highly sensitive national lab

Benefits

- Reduces time to capture and analyze 275MB of acoustic data from hours to one-fourteenth of a second
- Enables analysis of real-time data from different types of sensors and 1,024 individual channels to support extended perimeter security
- Enables a faster and more intelligent response to any threat



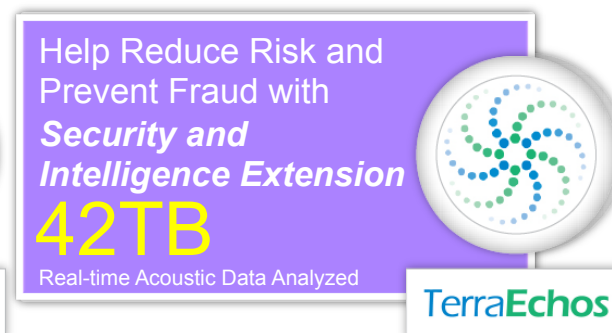
Harvest Business Value via Key Business-Driven Use Cases



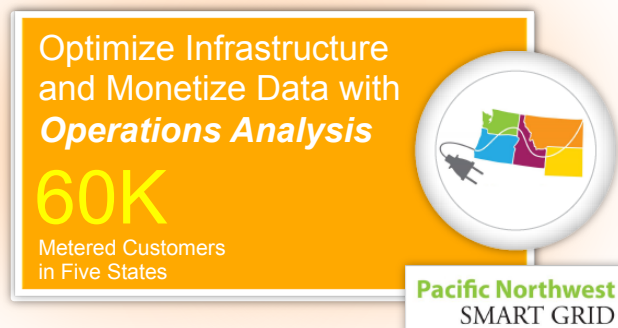
Big Data Exploration
Find, visualize, understand all big data to improve business knowledge



Enhanced 360° View of the Customer
Achieve a true unified view, incorporating internal and external sources



Security/Intelligence Extension
Lower risk, detect fraud and monitor cyber security in real-time



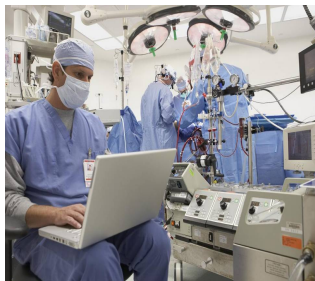
Operations Analysis
Analyze a variety of machine data for improved business results



Data Warehouse Augmentation
Integrate big data and data warehouse capabilities to increase operational efficiency

Big Data ... also can help to save costs, better understand customers, information systems, protect the environment, ...

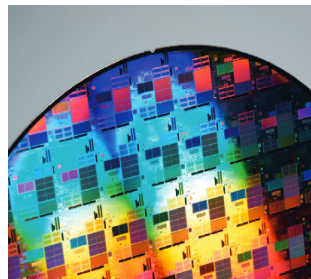
Medicine



Telecommunication



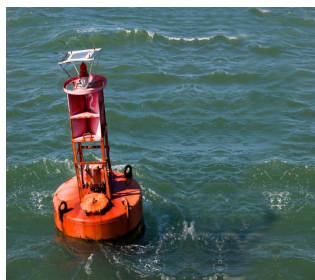
Manufacturing



Transport



Environment



Commerce



Healthcare



Radio astronomy



Industrie 4.0



Energy & Utilities



Connected Car



Big Data ... Knows you sometimes better than anybody else

TECH

2/16/2012 @ 11:02AM | 2,683,852 views

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

“ As Pole’s computers crawled through the data, he was able to identify about 25 products that, when analyzed together, allowed him to assign each shopper a “pregnancy prediction” score. More important, he could also estimate her due date to within a small window, so Target could send coupons timed to very specific stages of her pregnancy.

Source:

This

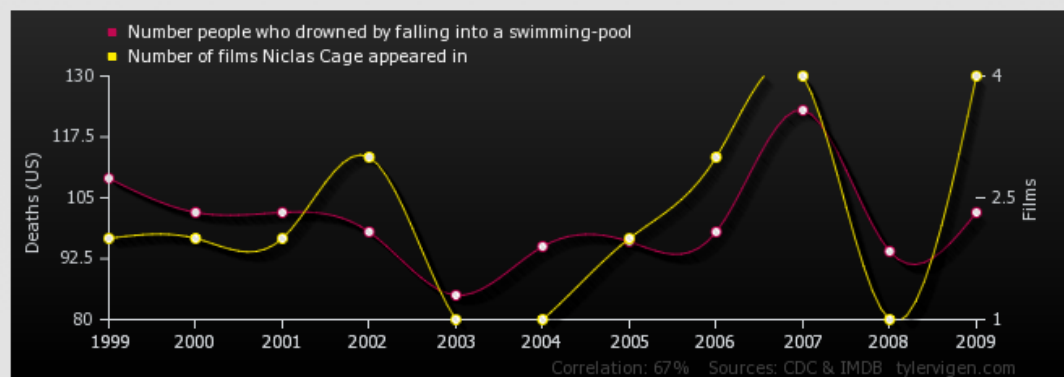
<http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/>

... But correlation $\neq$ causality, and correlations doesn't always make sense

Source:

http://tylervigen.com/view_correlation?id=359

Number people who drowned by falling into a swimming-pool
correlates with
Number of films Nicolas Cage appeared in



	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Number people who drowned by falling into a swimming-pool Deaths (US) (CDC)	109	102	102	98	85	95	96	98	123	94	102
Number of films Nicolas Cage appeared in Films (IMDB)	2	2	2	3	1	1	2	3	4	1	4
Correlation: 0.666004											

That's why the often cited data scientist requires skills in 3 main areas:
Technology, Mathematics and Business!

Analytics require an understanding of the context, not just data and algorithms!

Merging the Traditional and Big Data Approaches

Traditional Approach
Structured & Repeatable Analysis

Business Users
Determine what question to ask



IT
Structures the data to answer that question

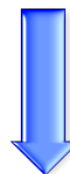


Monthly sales reports
Profitability analysis
Customer surveys

Big Data Approach
Iterative & Exploratory Analysis



IT
Delivers a platform to enable creative discovery



Business
Explores what questions could be asked



Brand sentiment
Product strategy
Maximum asset utilization



This is how data scientists usually work

Big Data is also about data security and ethical aspects

- **Anonymization/de-identification:**
replace attributes in a way that dependencies and relations between different data points are maintained but the personal information that identifies a person is removed, however, there is no 100% guarantee, sometimes people can be de-identified since the group of similar persons is very small
- **Opt-in vs. Opt-out**
- **Legal aspects**
- **Customer trust**
- **Experiments with data? And users?**

Value is often generated when using data in a complete new context ... such as search results to predict flu, but it can also be used for purposes that may not be obvious

Standortsuche für Radarfallen: TomTom entschuldigt sich für Deal mit der Polizei

Die Kunden sind wütend: In den Niederlanden hat der Navigationsgeräte-Hersteller TomTom erfasste Geschwindigkeitsdaten der Navi-Nutzer an die Polizei verkauft. Die will überall Radarfallen aufstellen, wo laut TomTom-Informationen statistisch viel gerast wird.

1 Donnerstag, 28.04.2011 – 12:53 Uhr

Drucken | Senden | Merken

Nutzungsrechte | Feedback

Kommentieren | 117 Kommentare

THEMA
Navigationssysteme

Alle Themenseiten

Mehr auf SPIEGEL ONLINE

Verhaltensüberwachung: Wie Handys die Welt beobachten (23.06.2010)

Mehr im Internet

Zeitung Algemeen Dagblad
in einem offenen Brief

SPIEGEL ONLINE ist nicht verantwortlich für die Inhalte externer Internetseiten.



REUTERS

Navigationgerät: TomTom setzt künftig eher auf den Verkauf von Verkehrsinformationen

Teilen | Empfehlen | 0 | Twittern | 4 | +1

Amsterdam - Der Navi-Hersteller TomTom hat die Kunden gegen sich aufgebracht. Das Unternehmen hat der Polizei in den Niederlanden die Geschwindigkeitsdaten ihrer Kunden verkauft. Man sei davon ausgegangen, dass die Daten benutzt werden, um den Verkehr auf den Straßen sicherer zu machen, schreibt TomTom-Chef Harold Goddijn in einem offenen Brief. Doch weit gefehlt: Nach Angaben der Zeitung "Algemeen Dagblad" plant die Polizei, anhand der gesammelten Geschwindigkeiten und den zugehörigen Ortsangaben geeignete Stellen für Radarfallen auszumachen. Wo viel gerast wird, sollen Radarfallen stehen.

Source:

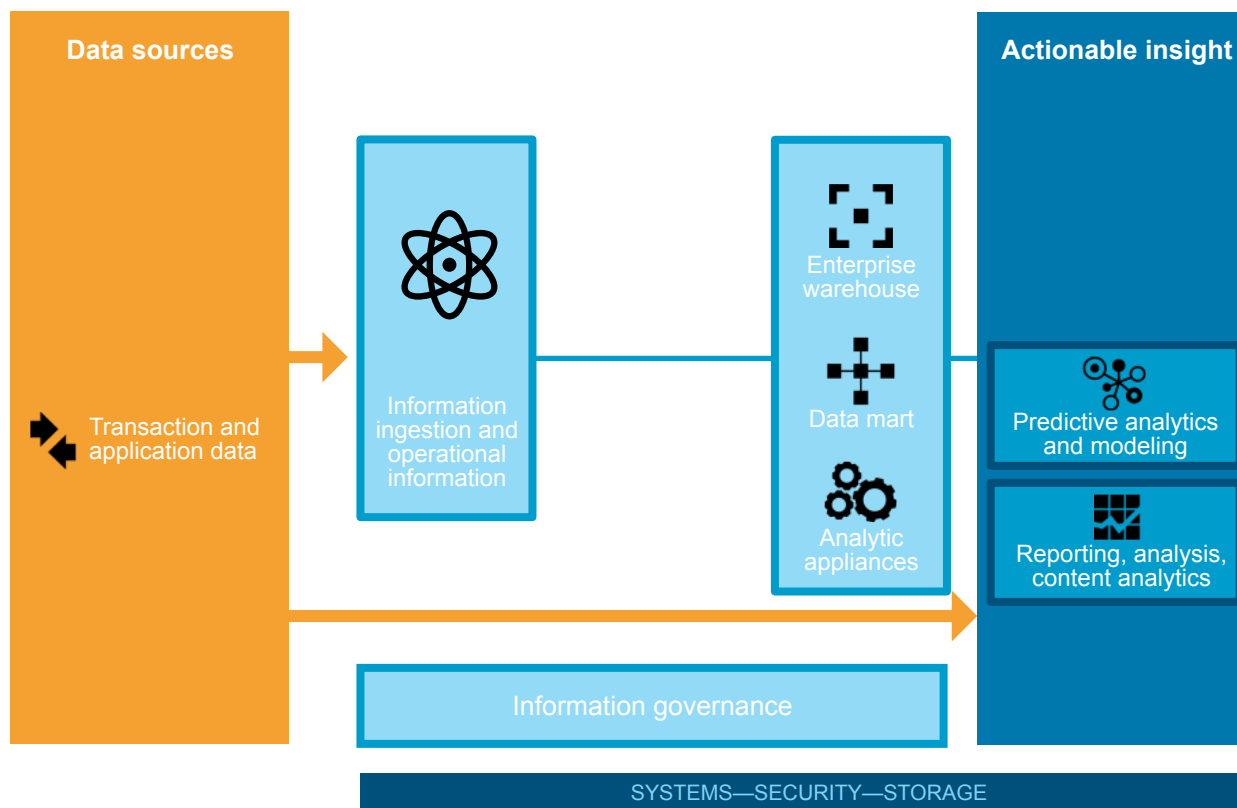
<http://www.spiegel.de/netzwelt/gadgets/standortsuche-fuer-radarfallen-tomtom-entschuldigt-sich-fuer-deal-mit-der-polizei-a-759464.html>

Big Data can answer some fundamental questions: What?, When? & Where?

- **What to offer** – the one thing everybody knows: “people who bought this also were interested in ...”
- But often you get plenty of recommendations, ... After you have bought a product, ... So using time considerations and **predictive** approaches probably makes sense to answer **when to offer what**
- Location based services offer a lot of opportunities to answer **where** to offer what, but are extremely security sensitive (since they permit total supervision) and haven't gained a lot of traction in Europe so far.
- It is all about **personalization**

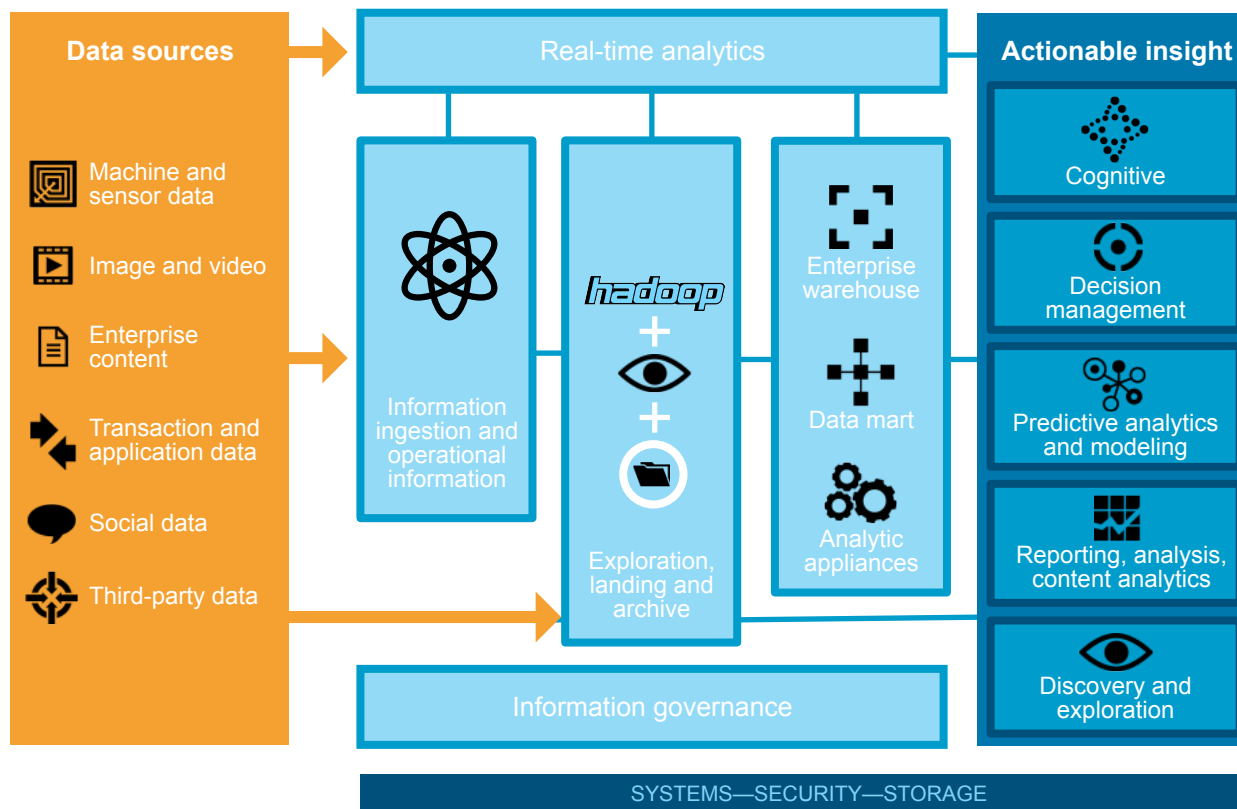
Next generation architecture

Starts from the current data management environment ...



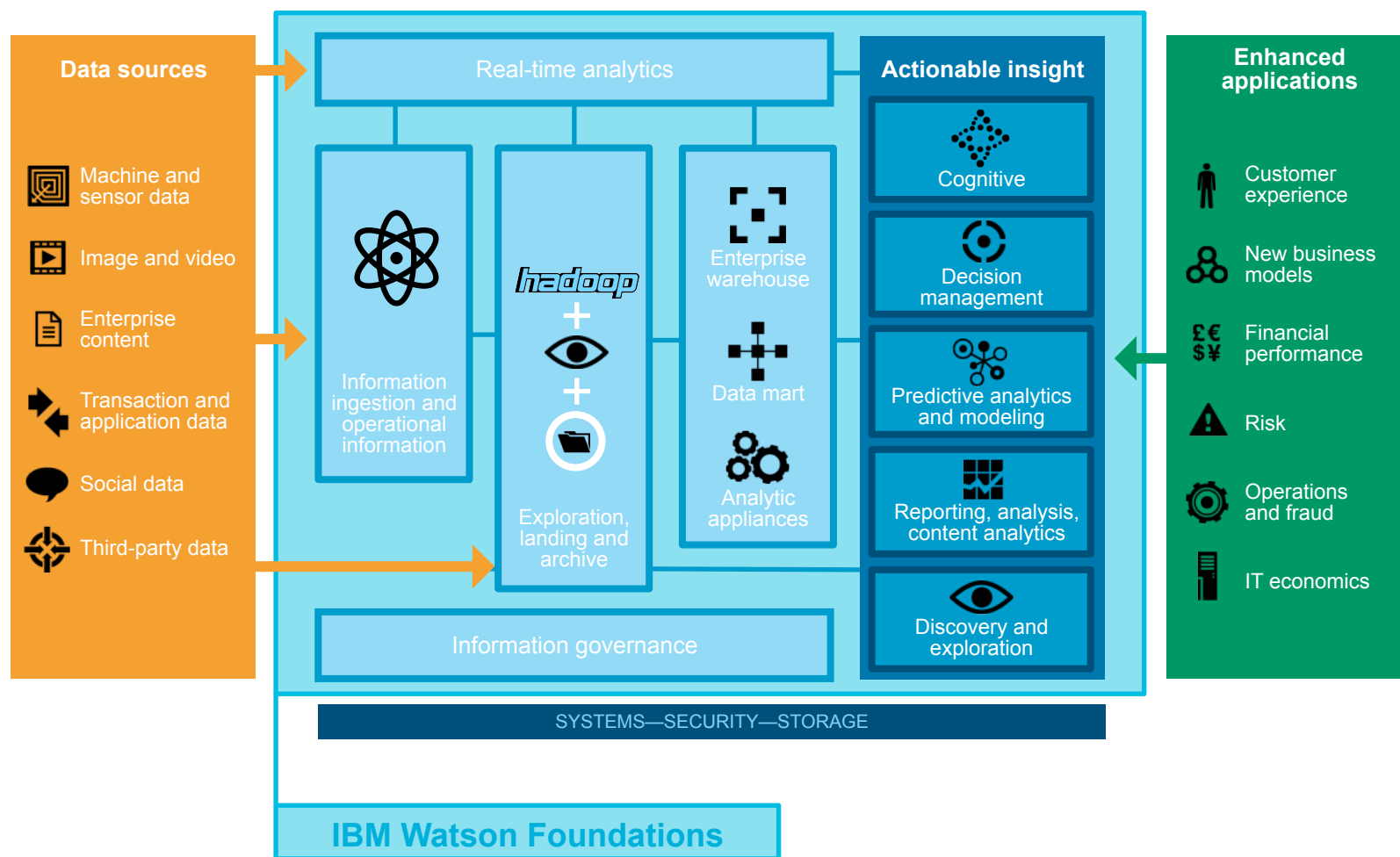
Next generation architecture

... adds new technologies and capabilities ...

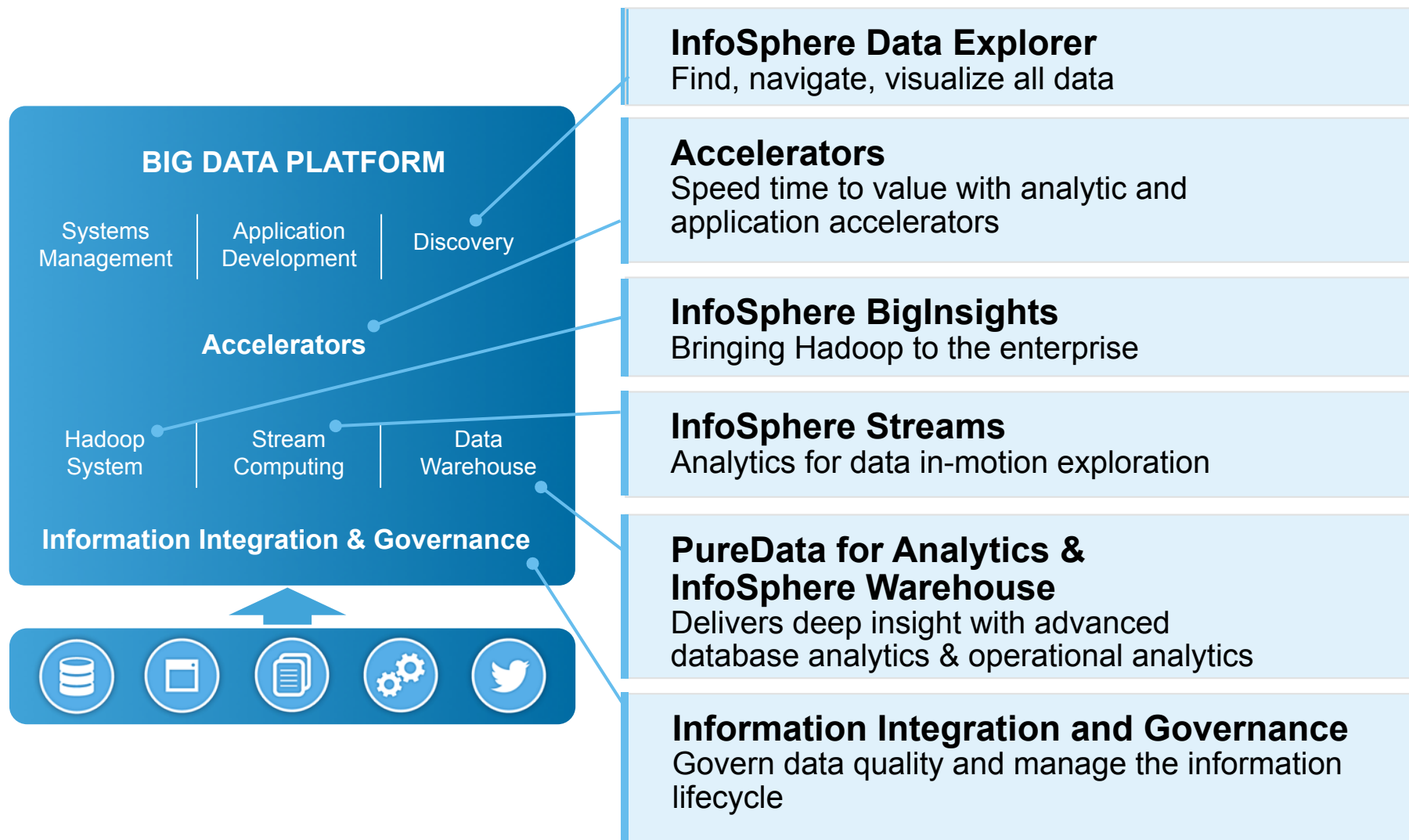


Next generation architecture

... to enable new applications



IBM's Key Platform Capabilities



Getting the Value from Big Data – Why a Platform?

The Whole is Greater than the Sum of the Parts

- Almost all big data use cases require an integrated set of big data technologies to address the business pain completely
- Reduce time and cost and provide quick ROI by leveraging pre-integrated components
- Provide both out of the box and standards-based services
- Start small with a single project and progress to others over your big data journey (*value is additive*)

Get Started on Your Big Data Journey Today

Get Educated

- IBM Big Data: ibm.com/bigdata
- IBMBigDataHub.com
- BigDataUniversity.com
- IBV study on big data
- Books / analyst papers

Schedule a Big Data Workshop

- Free of charge
- Best practices
- Industry use cases
- Business uses
- Business value assessment



THINK

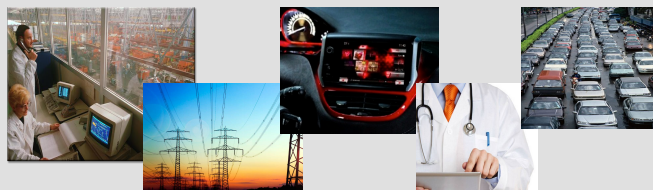
BIG

BIG

**Where technology meets business potential:
Start making sense of your data, the technology is ready for
prime time!**

**Do Marketing.
Be creative.
With Data.**

Gain value from your data



There are many opportunities to gain value from data. Let's talk how to make sense of your data!

IBM Big Data Workshop

Ein halber Tag, der sich voll auszahlt!



Holen Sie mehr für sich raus:
beim **IBM Big Data Workshop**

Alles drin. Alles kostenfrei. Jetzt Termin vereinbaren!

<http://www-05.ibm.com/de/events/workshop/bigdata/>