



# **Big Data: What You Should Know**

Mark Child

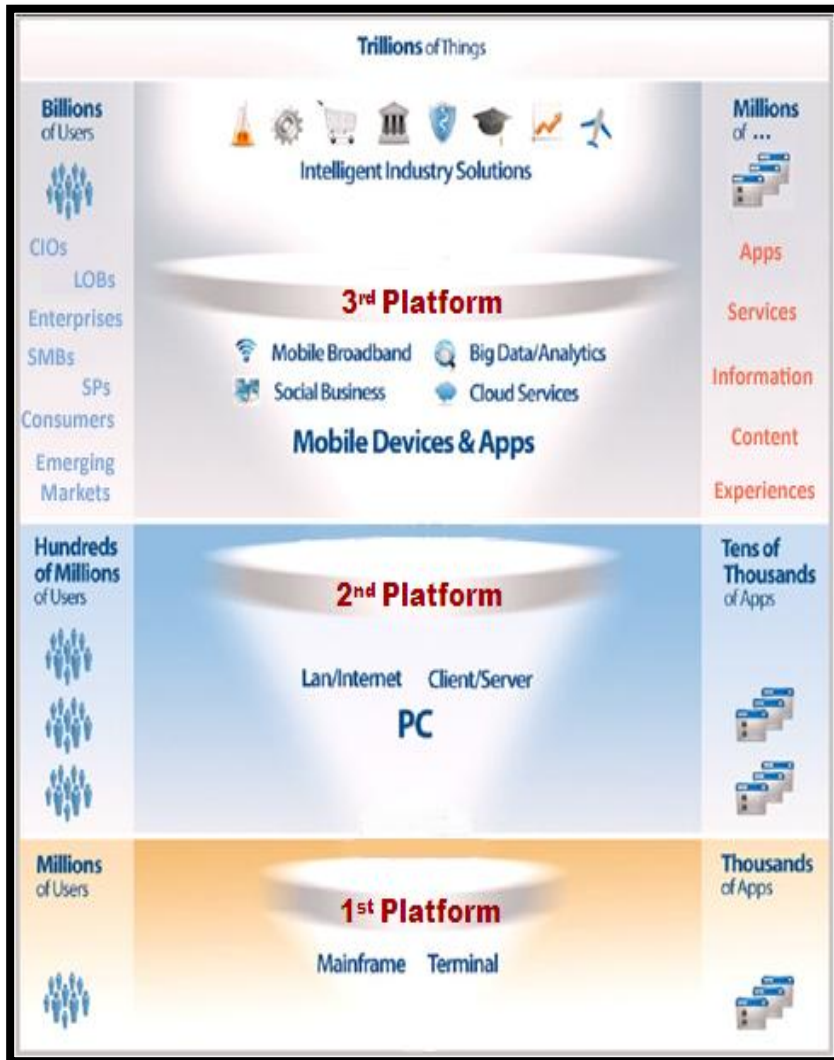
Research Manager - Software

IDC CEMA

# Agenda

- Market Dynamics
- Defining Big Data
- Technology Trends
- Information and Intelligence
- Market Realities
- Future Applications

# The IT Market and the 3<sup>rd</sup> Platform



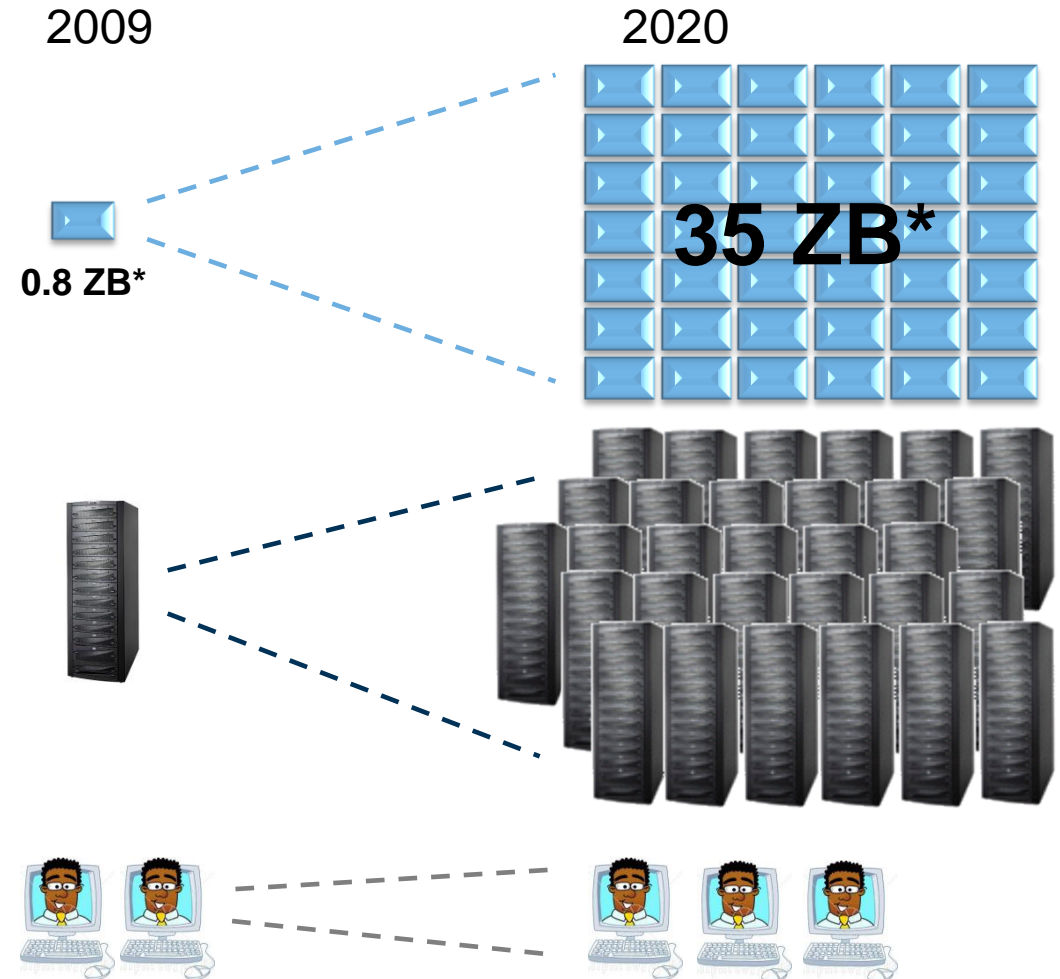
- The third platform is driven by four forces: **mobile broadband, devices and applications; social business; cloud services; and Big Data and analytics.**
- This platform revolves around:
  - **Millions of applications**
  - **Billions of users**
  - **Trillions of connected fixed and mobile devices**
- The second platform was that of **client-server** or where **PCs and servers were connected** through **LANs and the internet**, accompanied by tens of **thousands of applications** and **hundreds of millions of users.**
- The first platform was an industry characterized by **mainframe computers** and **terminals** with **thousands of applications** being delivered to **millions of users.**

# Data Growing Unabated

- Information will grow by factor of 44 from 2009 to 2020

- Storage capacity will grow by a factor of 30

- IT staffing will grow by a factor of 1.4



\*Zettabyte = 1 trillion gigabytes

# Sources of Data Growth

Office Apps



Enterprise Apps



Web Apps



Digital security



**Data creation  
is part of our  
daily lives**

Connected devices  
and media



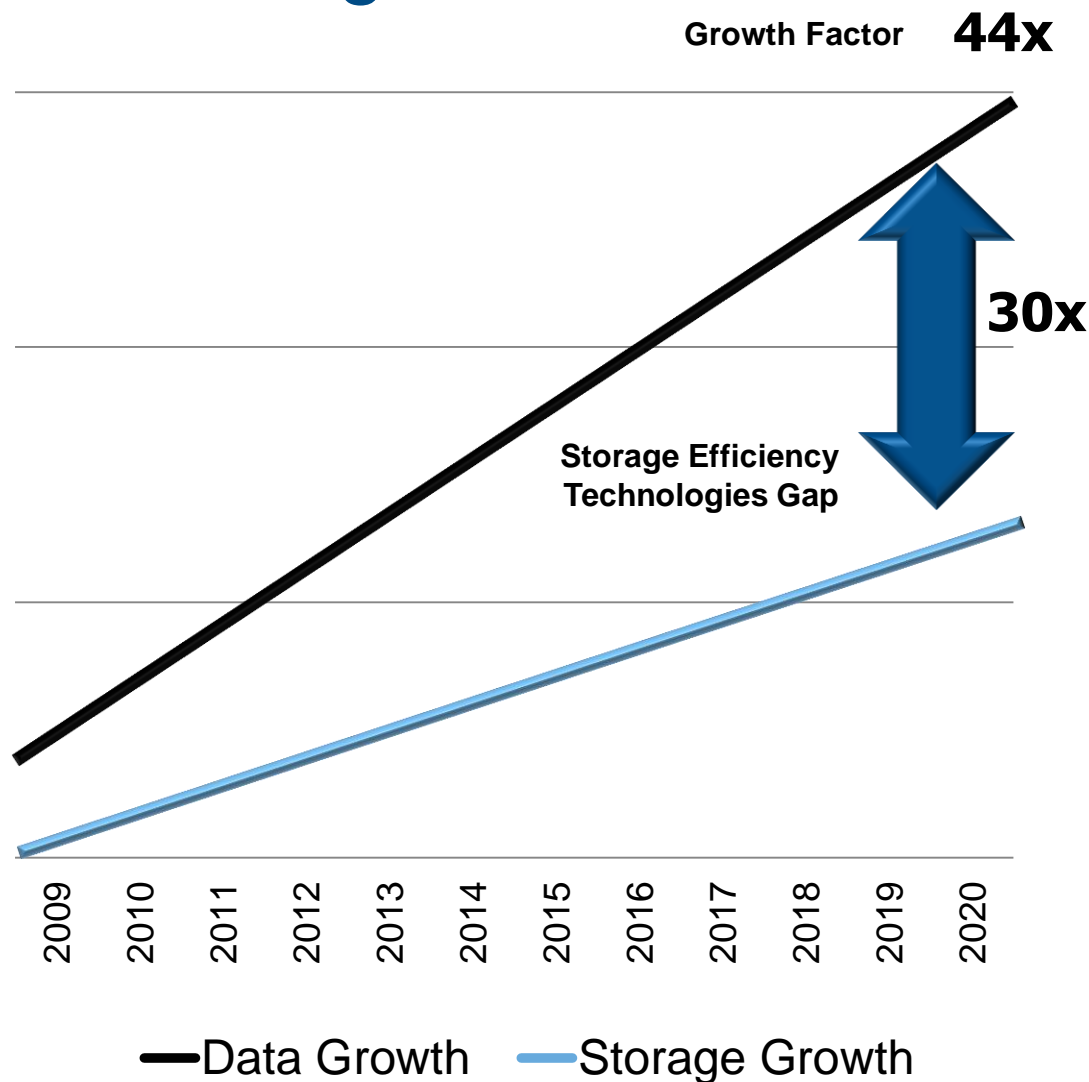
Social networks



Software as a Service



# Acceleration in the Adoption of Storage-Efficient Technologies



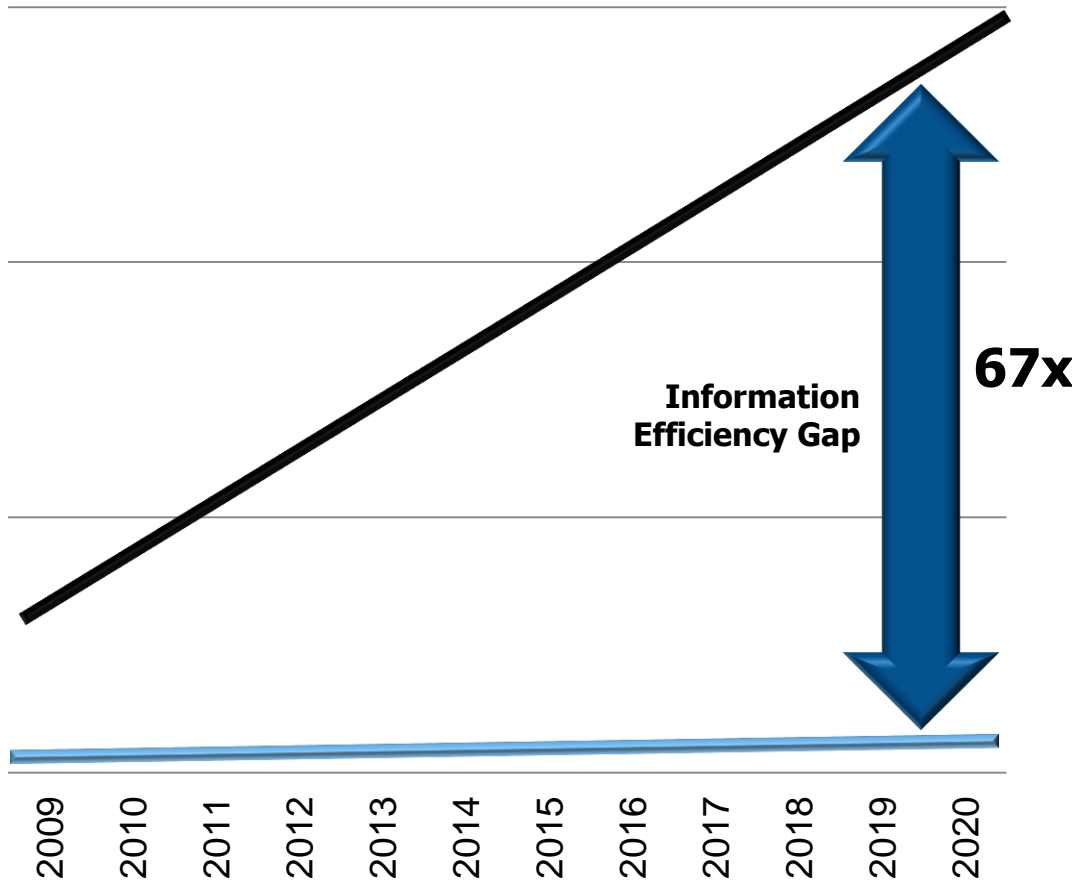
## Storage Efficiency

- **Storage Virtualization** - CAPEX, Migration costs, Management overhead, Scalability, Uptime
- **Caching/SSD** - SLA/QoS, Floorspace, Power
- **Thin Provisioning** - Reclaim unused capacity
- **Tiering** - Raw cost / per capacity
- **Deduplication, Compression** - Get rid of copies, Data footprint
- **Unified/Converged Mgmt.** - Management overhead
- **Hypervisor integration** - VM provisioning overhead
- **Integrated snapshots and replication** - DP & DR overhead
- **FCoE, iSCSI** - Storage network costs



# Acceleration in the Adoption of Information-Efficient Technologies

Growth Factor



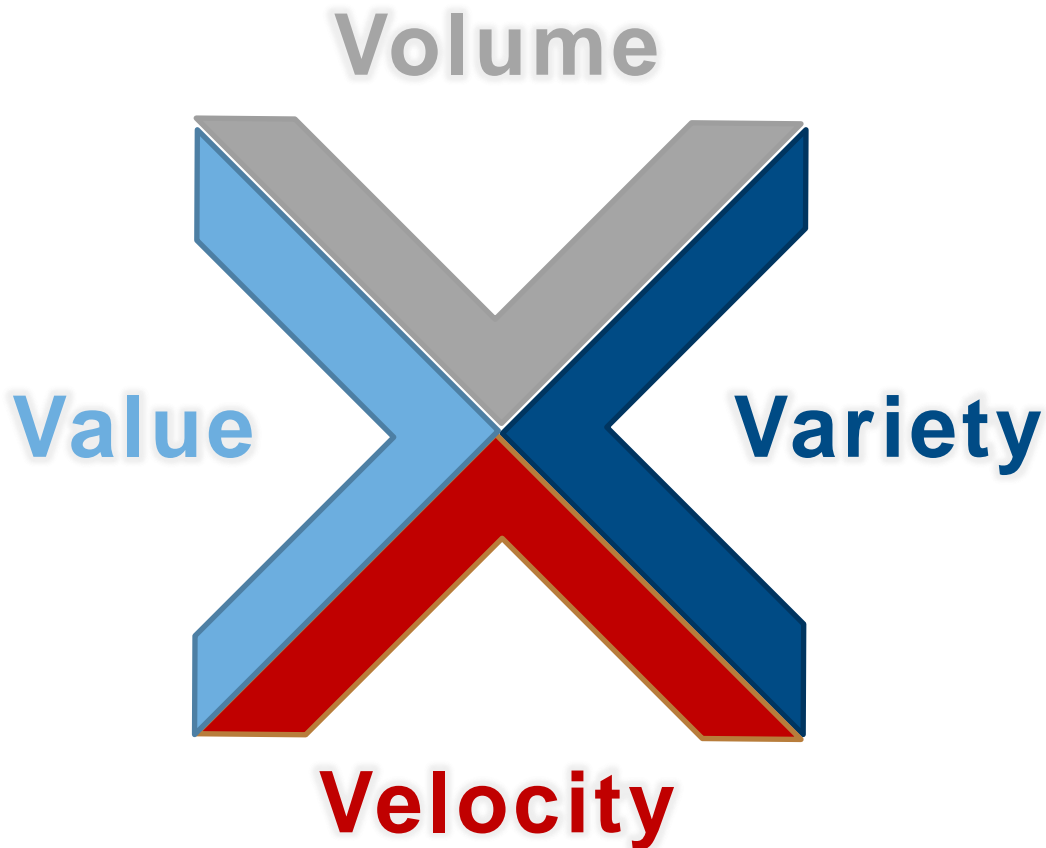
## Information Efficiency

- Advanced Analytics Software (Embedded and Stand-alone)
- EQRA – End-user Query, Reporting and Analysis
- Data Warehouse Generation and Management
- Content Analysis Tools
- Hadoop, Hbase, Map reduce

— Data Objects Growth

— Staffing Growth

# What is Big Data?



IDC defines Big Data technologies as a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high velocity capture, discovery and/or analysis.



# How does Big Data differ from Business Analytics?

Greater volumes of data than ever before  
Volumes growing quickly

**Volume**

Cost reduction –  
preventing  
system and  
manufacturing  
faults

Revenue  
generation –  
preventing online  
store outages

**Value**

**Variety**

Not just RDBMS  
data but  
unstructured  
information  
e.g.  
System logs  
Social information

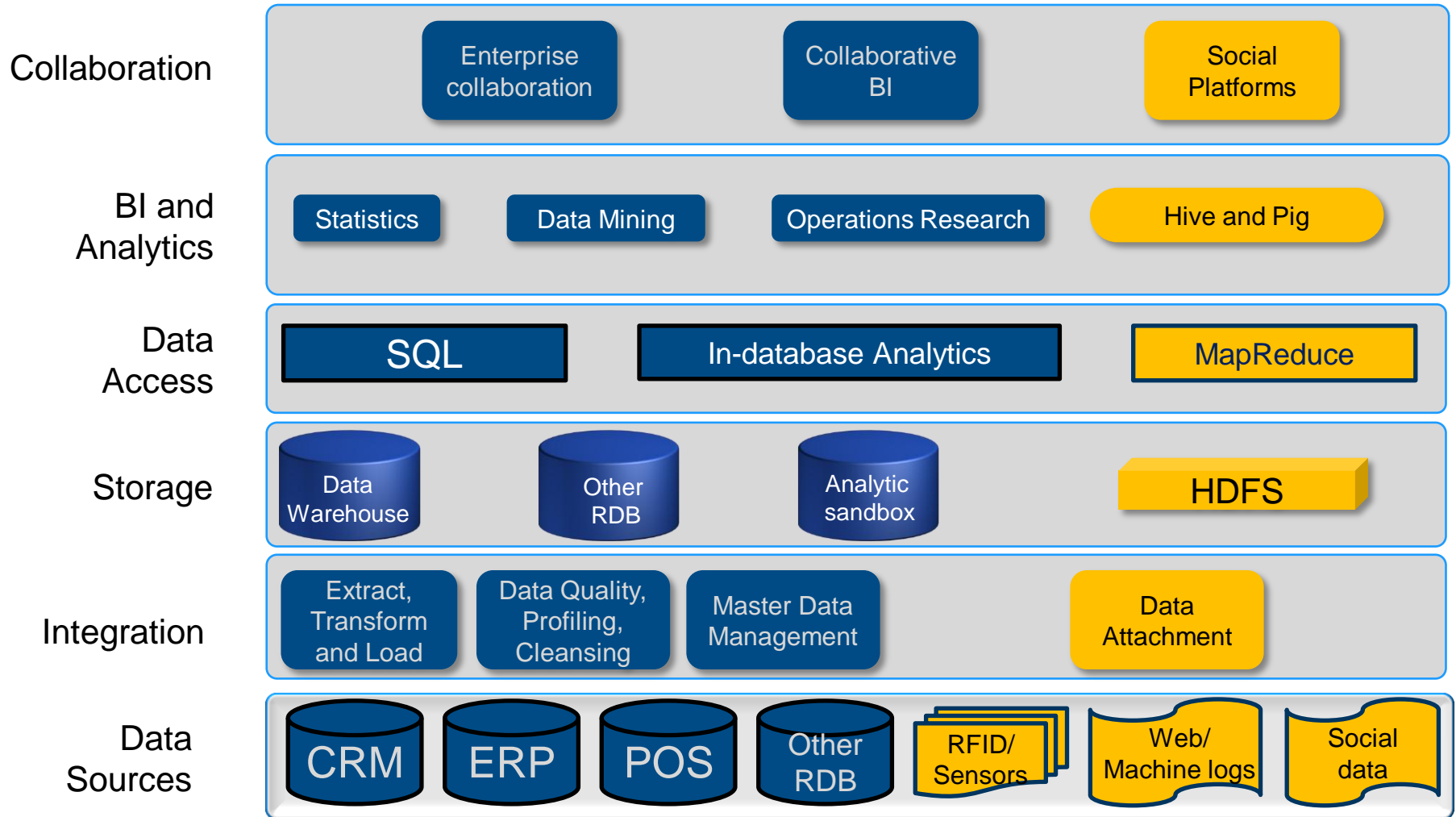
**Velocity**

Data processing in near-real-time  
(e.g. processing system logs to predict and prevent faults)  
Sensor data

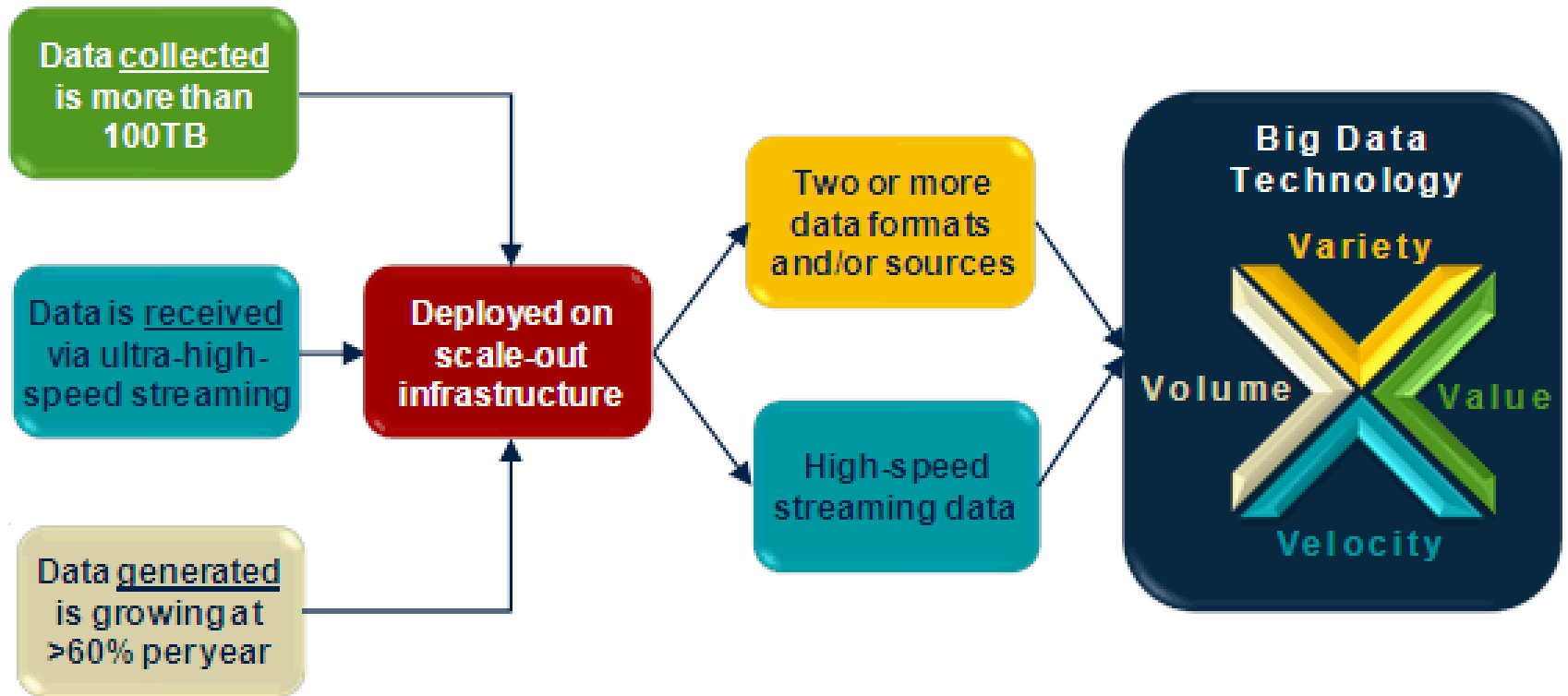
# Data Source – Big Data Output

Process-mediated data	Machine-generated data	Human-sourced information
Transactional Tabular Relational Metadata Operational Structured Managed Regulated System	Sensor data Computer logs Structured Reliable Fast High volume “The Internet of Things”	Social Multimedia Loosely structured or unstructured Unregulated Subjective Insightful Unreliable? Requires standardization

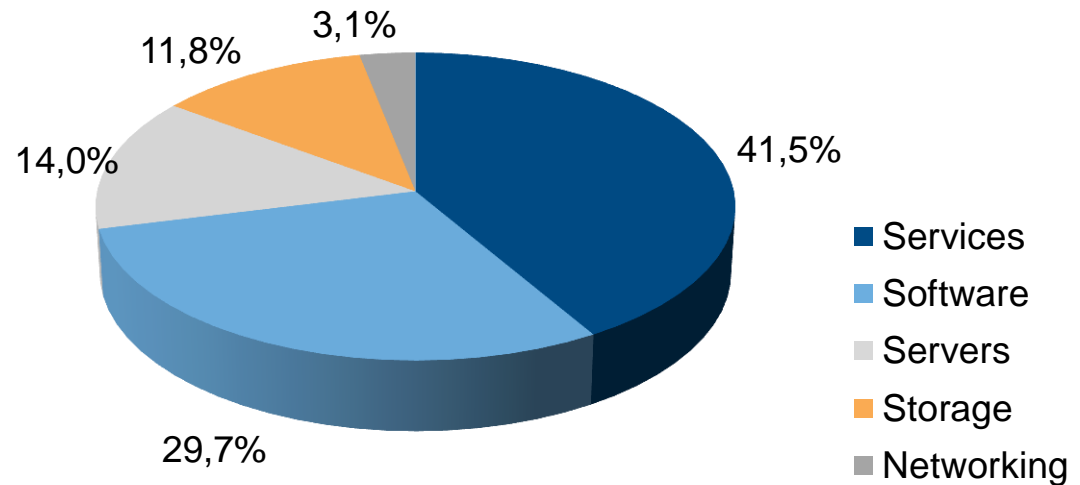
# Is This Big Data?



# IDC's Big Data Technology and Services Market Sizing Criteria



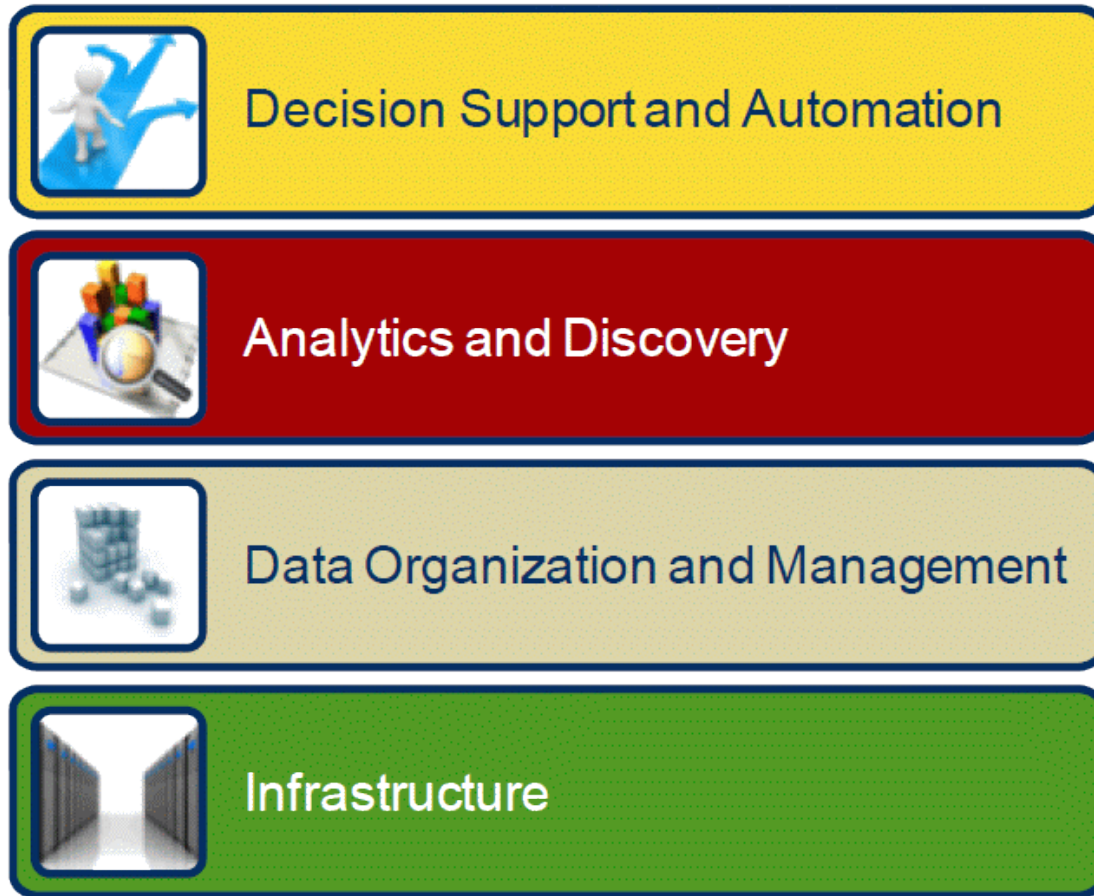
# Worldwide Big Data Technology and Services Revenue Share by Segment, 2011



Total: \$4.8 billion

Source: IDC, 2012

# The Big Data technology stack



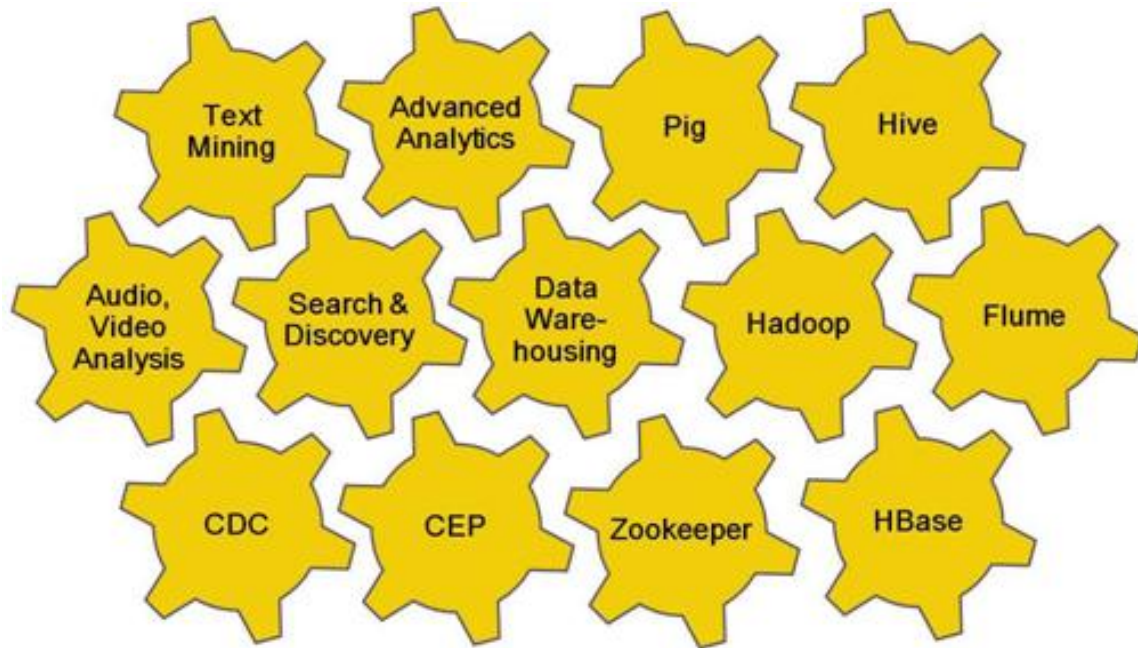
Source: IDC, 2012

# Big Data and Analytics:

## The Chief CIO Issue from 2013 Onwards

- CIOs are being overwhelmed by the amount of data they are asked to manage
- Big Data and Analytics will be the issue of the year for many CIOs in 2013 in an effort to provide more value from IT
- CIOs will conduct numerous pilot analytics projects – not all will succeed – until they find the right tools and data models to provide best value

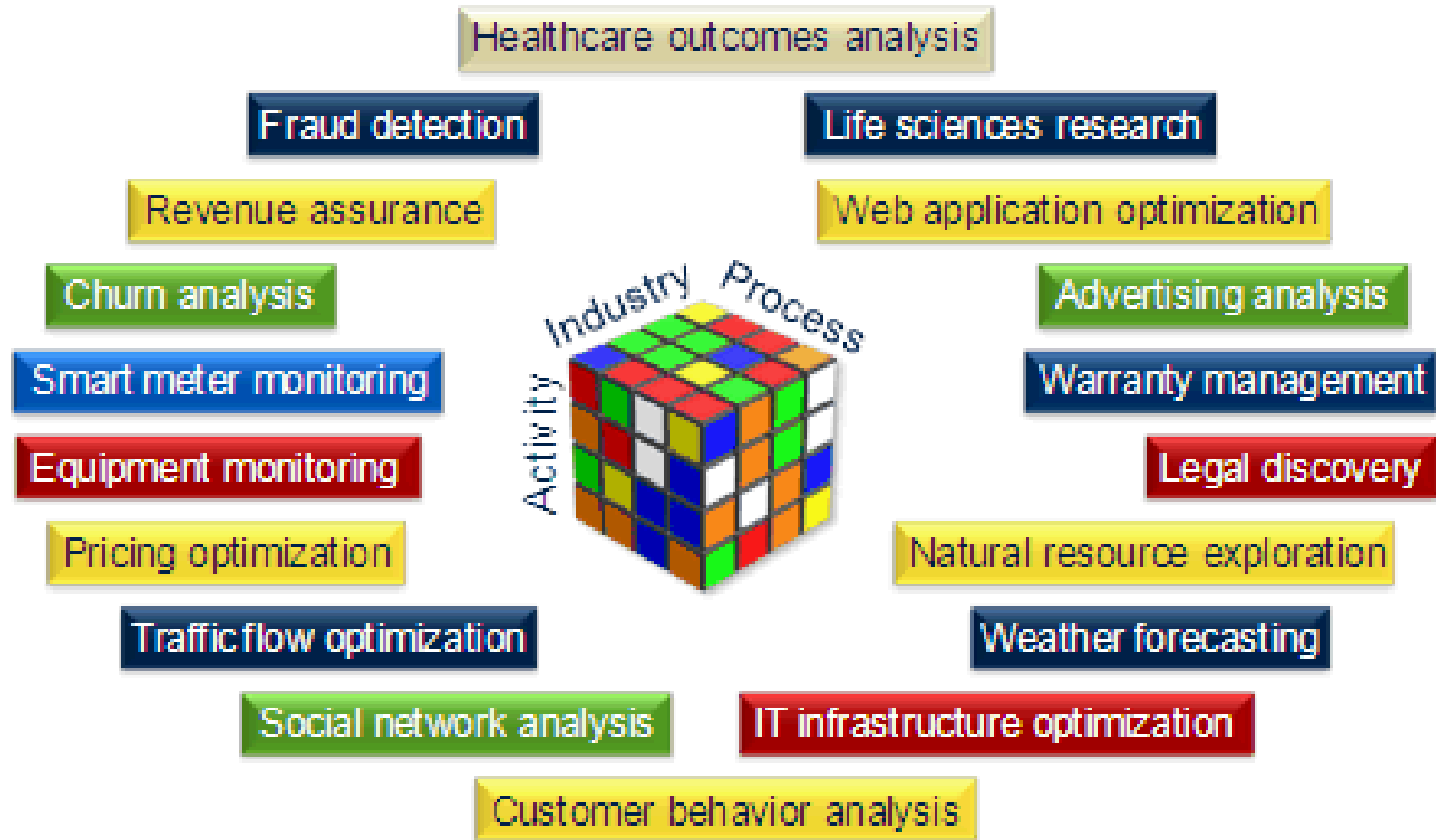
### Big Data Software Landscape



Source: IDC, 2013



# Big Data technology and services use cases



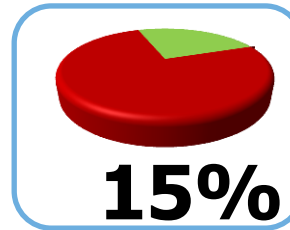
# The Value Of Information – Where To Find It



# Data Usage

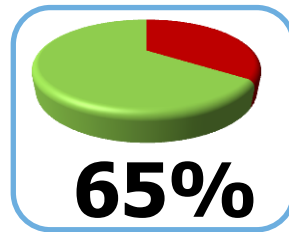
## *Being Practical*

Out of 100% of data saved,  
only



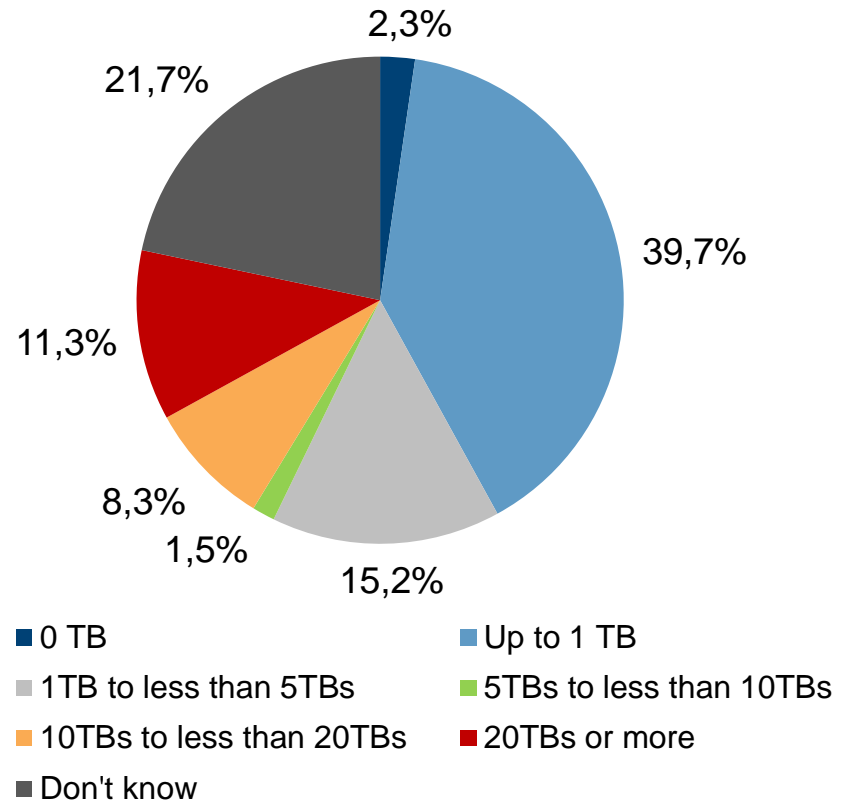
is accessed in future.

Out of 15%



is accessed only once.

# How much data does your organization currently process per day?



Source: IDC IT Buyers Pulse 2012

N=566 companies across CZ, HU, PL, RO, RU

# What are the key areas that benefit (or would benefit) the most from leveraging Big Data solutions in your organization?

Source: IDC IT Buyers Pulse 2012

N=425 companies across CZ, HU, PL, RO, RU



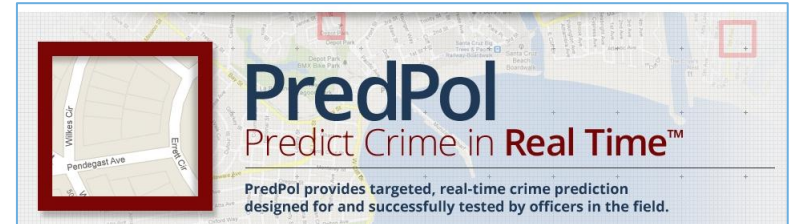
# Big Data and Intelligent Security



- Monitoring and logging, SIEM help enable shift from reactive to proactive security measures
- Big Data's ability to process large volumes of raw data in real time can enable shift from proactive to predictive security:
  - Pattern recognition and predictive analytics
  - Link to automated defensive systems: network traffic blocking, system quarantine, extended identity verification
  - Risk analysis and risk management
  - Integration with GRC systems

# Big future for big data?

**SMART DATA**



**NewSQL**

**Data Science**

# Predictive Analytics



# Challenges

- Personnel and skills
- Data governance
- Sponsorship
- Data classification, lifecycle and flow
- Implementing business process changes

# Thank you!

Mark Child  
Research Manager - Software  
IDC CEMA  
[mchild@idc.com](mailto:mchild@idc.com)  
221 423 140

