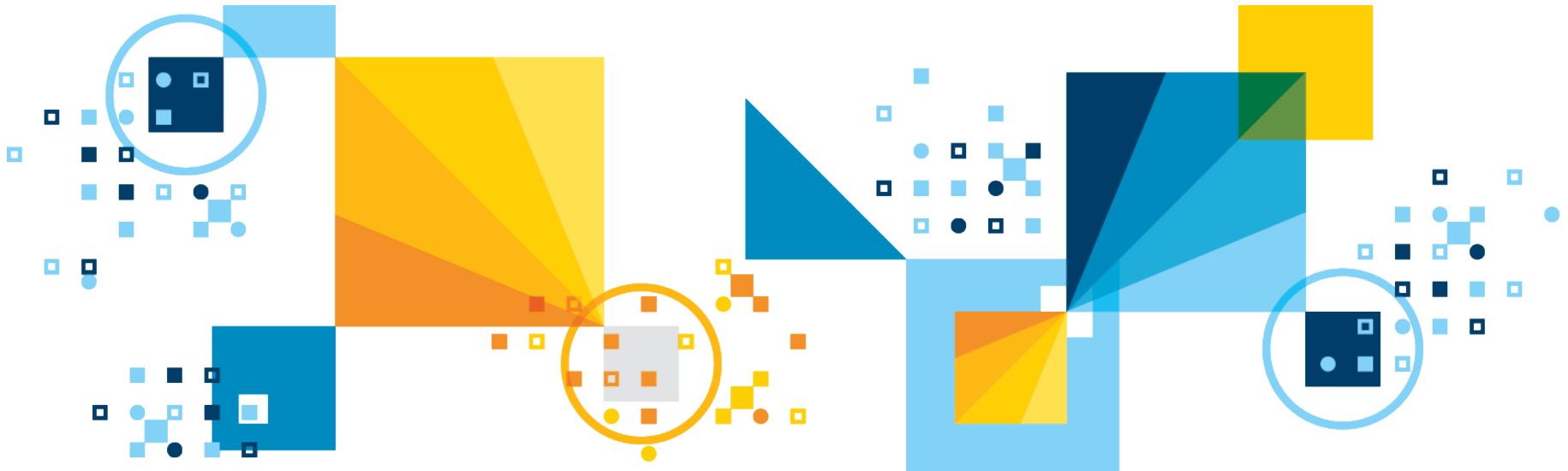
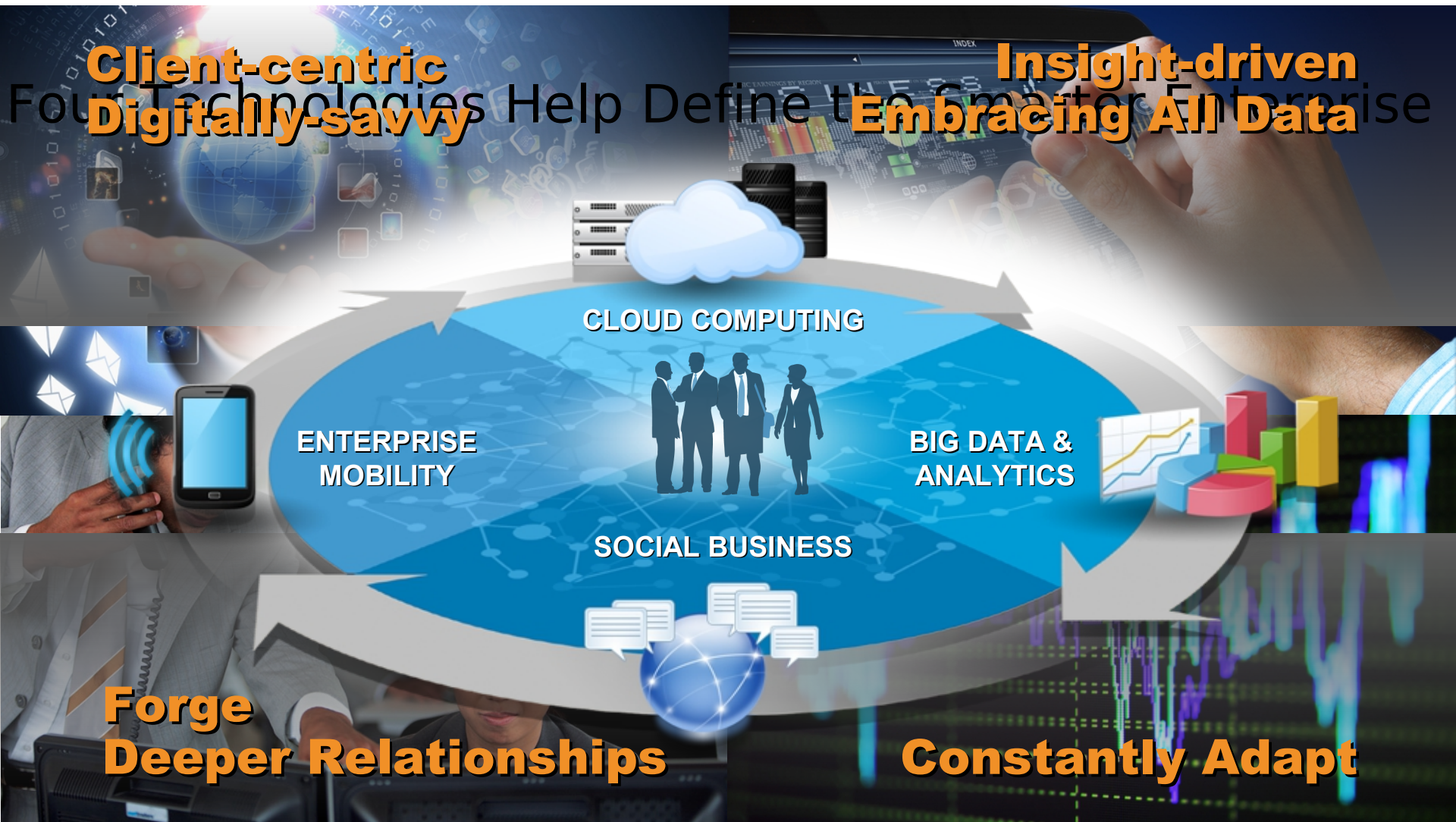


# Big Data:

## Modern Ecosystems for Data Warehousing and Analytics

*James Kobiulus*  
*IBM Big Data Evangelist*





# The World of Big Data & Analytics Is Rapidly Expanding



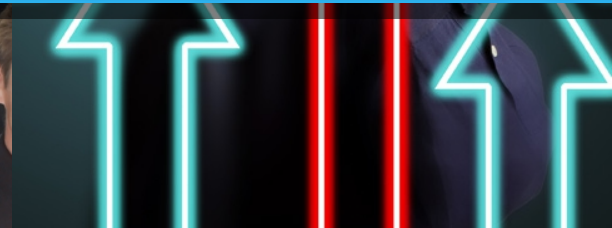
**Untapped Resource**



**Empower  
Everyone**



**Increased Value**



# We Have Entered a New Era of Computing . . .

Tabulating  
Systems Era



1900

Programmable  
Systems Era



1950

Cognitive  
Systems Era



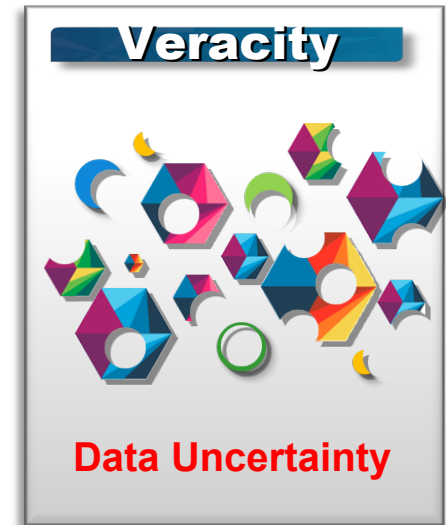
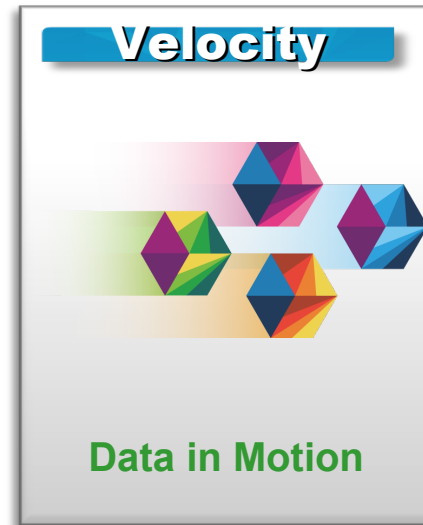
cog·ni·tive: of or pertaining to the mental processes of perception, memory, judgment, learning, and reasoning.

2011

. . .Enabling New Opportunities and Outcomes



# Big Data Is All Data



# What is Big Data?

*Extracting business insight from advanced analytics at extreme scale*

## Volume

**12** terabytes  
of Tweets create daily

Analyze sentiment

## Velocity

**5** million  
trade events per second

Identify potential fraud

## Variety

**100's** video  
feeds  
from surveillance cameras

Monitor events of interest

**350** billion  
meter readings per annum

Predict power consumption

**500** million  
call detail records per day

Prevent churn

**80%** data  
growth  
are images, video, documents...

Improve customer satisfaction

# What is Big Data NOT?

## True or False

*Big Data is only about unstructured information.*

**False!** Most projects include structured information sources.

*Big Data technologies require huge amounts of data.*

**False!** Flexibility, not data size, is the most important aspect.

*Big Data projects are expensive.*

**False!** You should start small and grow fast.

*Big Data projects require a “big” idea.*

**False!** Most projects are about making data more readily available, not changing the world.

*Big Data = Apache Hadoop*

**False!** Big data technologies are numerous and diverse and need to be matched to your problem.

# Identify Core Big-Data Use Cases in Your Operations



## Big Data Exploration

Find, visualize, understand all big data to improve decision making



## Enhanced 360° View of the Customer

Extend existing customer views (MDM, CRM, etc) by incorporating additional internal and external information 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

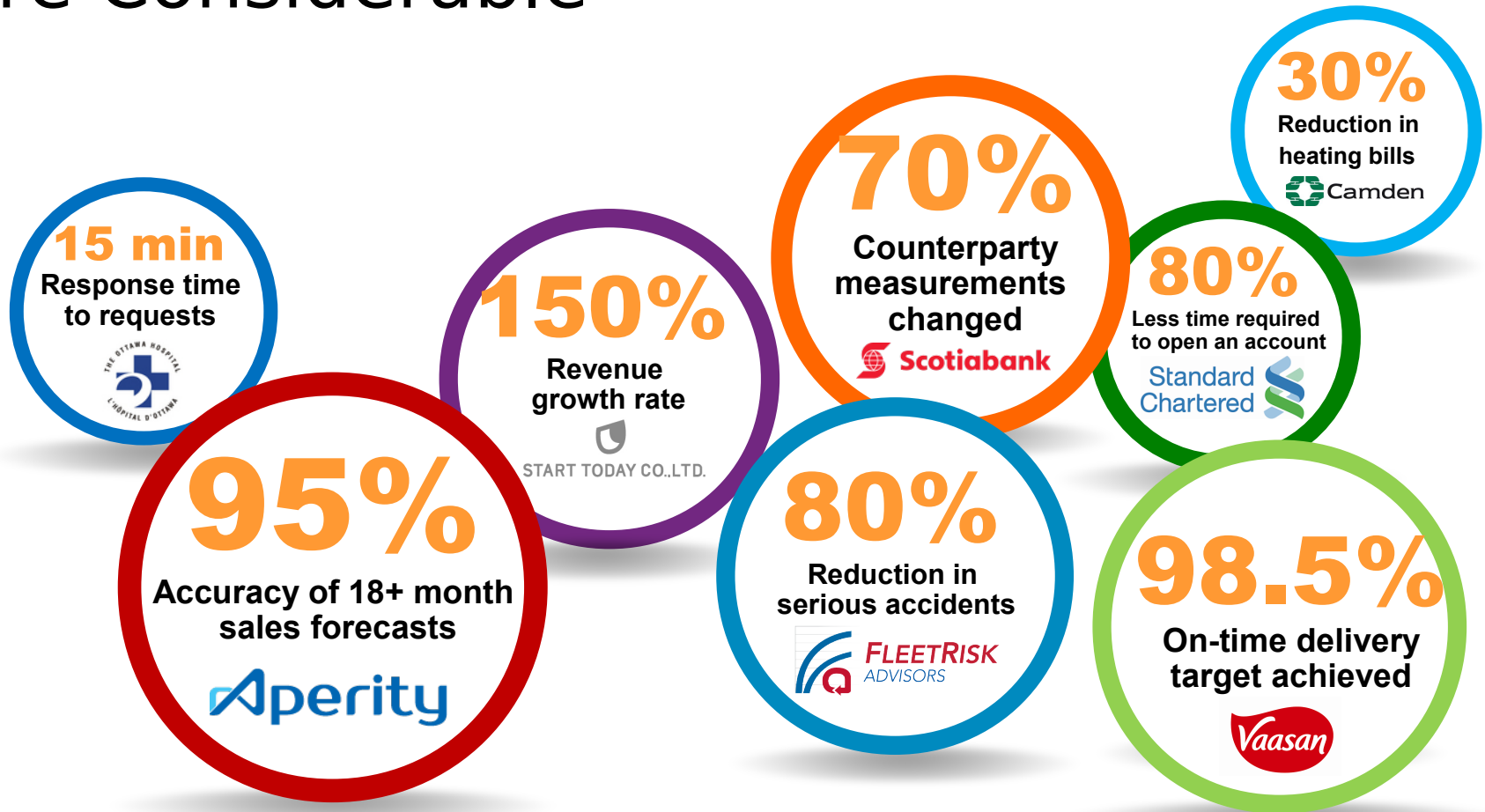


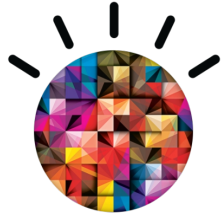
## Data Warehouse Augmentation

Integrate big data and data warehouse capabilities to increase operational efficiency



# The Returns from Big Data & Analytics Are Considerable





# Three Key Imperatives for Big Data & Analytics Success

Be proactive  
about privacy,  
security and  
governance

**Trust It.**

Build a culture  
that infuses  
analytics  
everywhere

**Imagine It.**

Invest in a  
big data &  
analytics  
platform

**Realize It.**

# **Trust It.** Be proactive about privacy, security and governance.

## **Trust the facts**



**Create foundation  
of trusted data**

## **Ensure privacy and security**



**Understand usage and  
monitor compliance**

## **Make risk aware decisions**



**Model exposure and  
understand variability**

***And trust the platforms, providers, and  
ecosystems you've incorporated into your  
big-data analytics initiatives.***

# Trust It. Don't Skimp on Data Quality, Protection, & Governance

## Integrate & Link Big Data

- High performance load and read
- Integrate seamlessly w/existing enterprise
- Data lineage & impact analysis across the data supply chain



## Cleanse and Validate Big Data

- Accuracy and entity matching with social data and standardization of machine data
- Cleanse and match final results of big data analysis before reporting



## Protect Big Data

- Active monitor trap any unexpected access and breaches
- Mask sensitive data to ensure data protection and industry compliance



## Master Big Data

- Manage and consume master data in all downstream analytics
- Link big data to trusted golden records
- Entity resolution at extreme scale



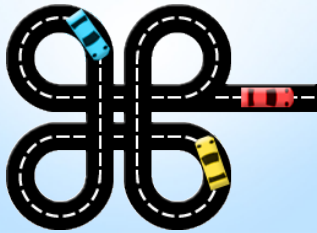
## Audit & Archive Big Data

- Create queryable archive for exploratory analysis of historical data
- Ensure compliance with immutable archive and secure access to the data



# Imagine It. *Extracting insights from all of your data.*

**Real-time  
Traffic Flow  
Optimization**



**Fraud & risk  
detection**



**Understand and act  
on customer  
sentiment**



**Accurate and timely  
threat detection**



**Predict and act on  
intent to purchase**

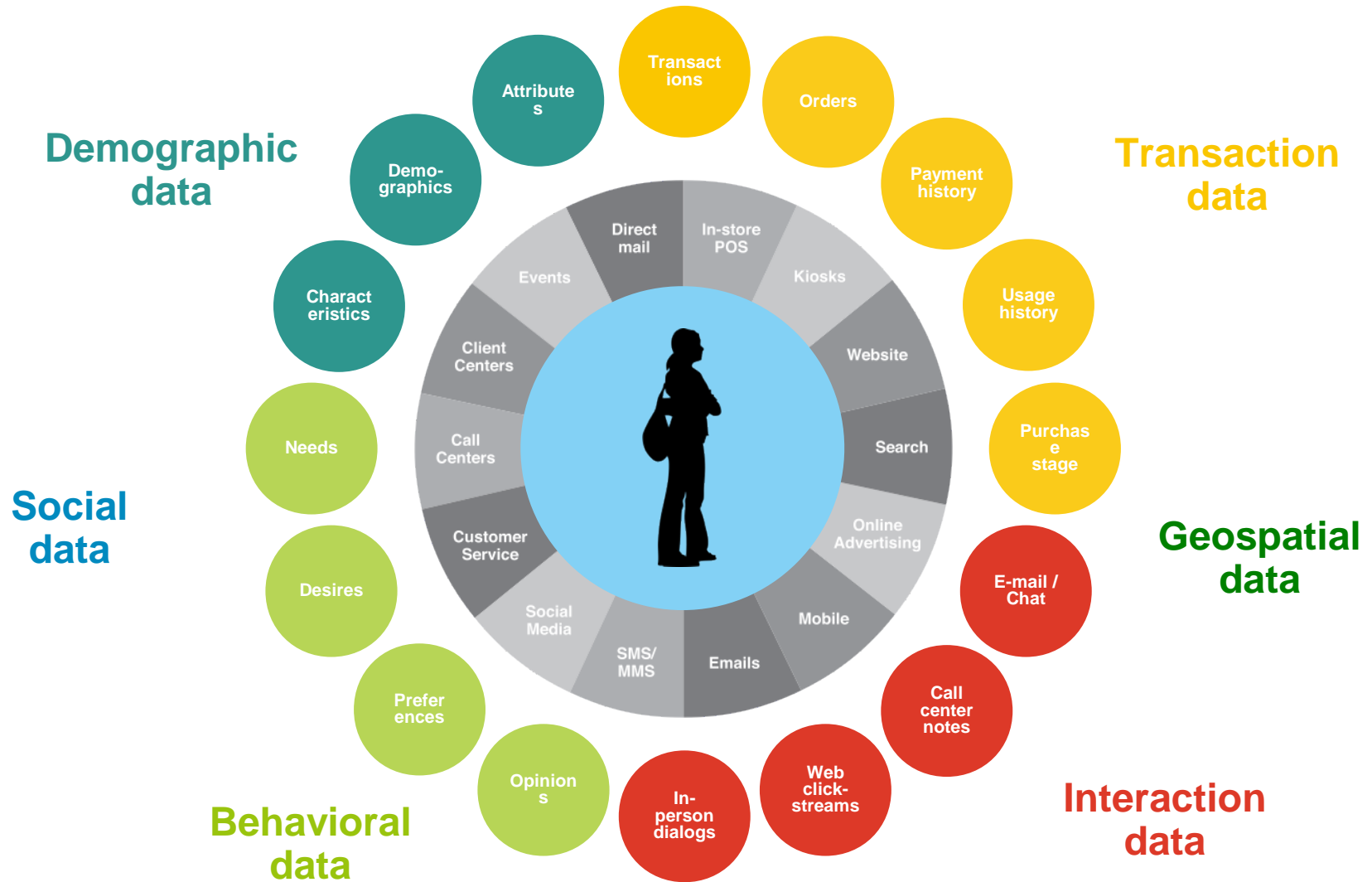


**Low-latency  
network analysis**

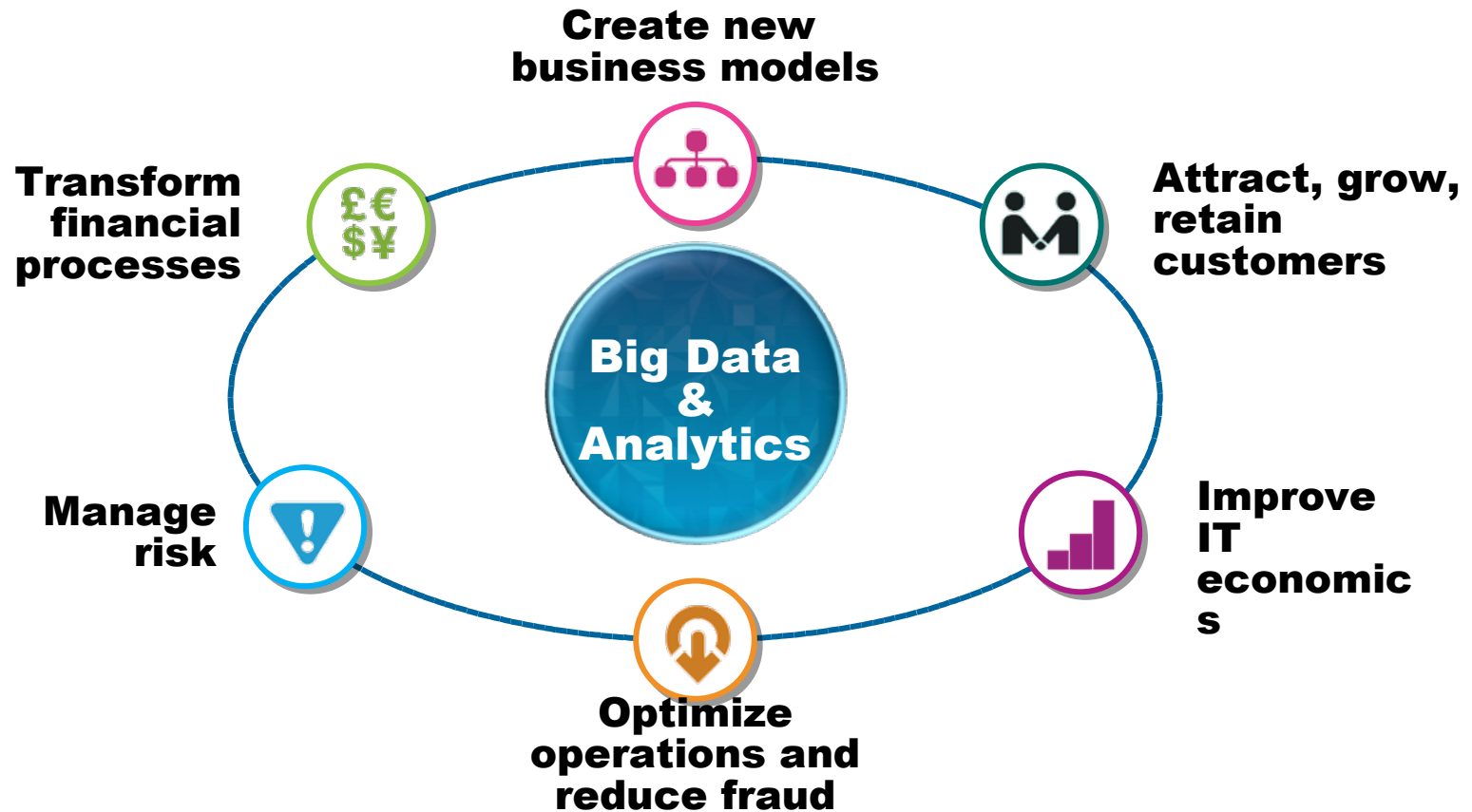




# Imagine It. *Unlock insights by merging internal & external data.*



# Imagine It. Infuse analytics into key business processes.



# Deliver Data-Driven Insights Throughout the Enterprise

## Administrators

...secure, manage, and optimize data access and analysis operations

## Business Users

...offer personalized price promotions to different customer segments in real-time

## Business Development

... find and deliver new mechanisms to monetize network traffic and partner with upstream content providers

## Data Scientists

... analyze subscriber usage pattern in real-time and combine that with the profile for delivering promotional or retention offers

## Business Executives

...get real-time reports and analysis based on data inside as well as outside the enterprise (web, social media etc.)

## Business Analysts

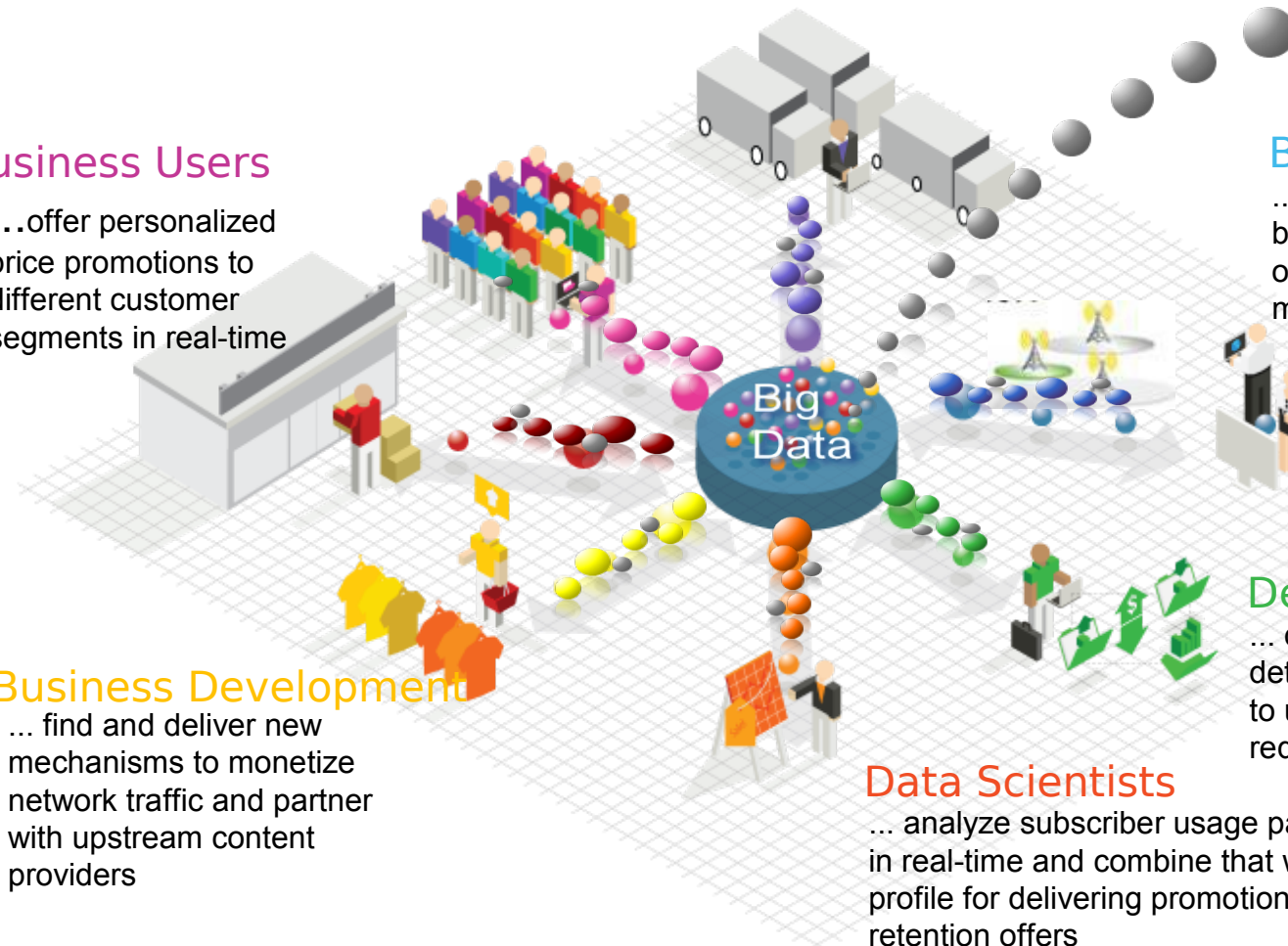
... analyze social media buzz for the new services/offerings to gauge initial success and any course correction needed

## Developers

... develop new apps and detailed algorithms in response to user and business requirements



External Data



# Imagine It. Every industry can leverage big data and analytics.



## Banking

- Optimizing Offers and Cross-sell
- Customer Service and Call Center Efficiency



## Insurance

- 360° View of Domain or Subject
- Catastrophe Modeling
- Fraud & Abuse



## Telco

- Pro-active Call Center
- Network Analytics
- Location Based Services



## Energy & Utilities

- Smart Meter Analytics
- Distribution Load Forecasting/Scheduling
- Condition Based Maintenance



## Media & Entertainment

- Business process transformation
- Audience & Marketing Optimization



## Retail

- Actionable Customer Insight
- Merchandise Optimization
- Dynamic Pricing



## Travel & Transport

- Customer Analytics & Loyalty Marketing
- Predictive Maintenance Analytics



## Consumer Products

- Shelf Availability
- Promotional Spend Optimization
- Merchandising Compliance



## Government

- Civilian Services
- Defense & Intelligence
- Tax & Treasury Services



## Healthcare

- Measure & Act on Population Health Outcomes
- Engage Consumers in their Healthcare



## Automotive

- Advanced Condition Monitoring
- Data Warehouse Optimization



## Chemical & Petroleum

- Operational Surveillance, Analysis & Optimization
- Data Warehouse Consolidation, Integration & Augmentation



## Aerospace & Defense

- Uniform Information Access Platform
- Data Warehouse Optimization



## Electronics

- Customer/ Channel Analytics
- Advanced Condition Monitoring



## Life Sciences

- Increase visibility into drug safety and effectiveness

**Realize it.** Develop your people's analytical skillsets.

**20%**

U.S. Department of  
Labor Forecasts\*



\*More than 20 percent increase in analytics-based jobs between now and 2018.



# Realize It. The spectrum of analytics is expanding.

**Be  
More  
Right,  
More Often**

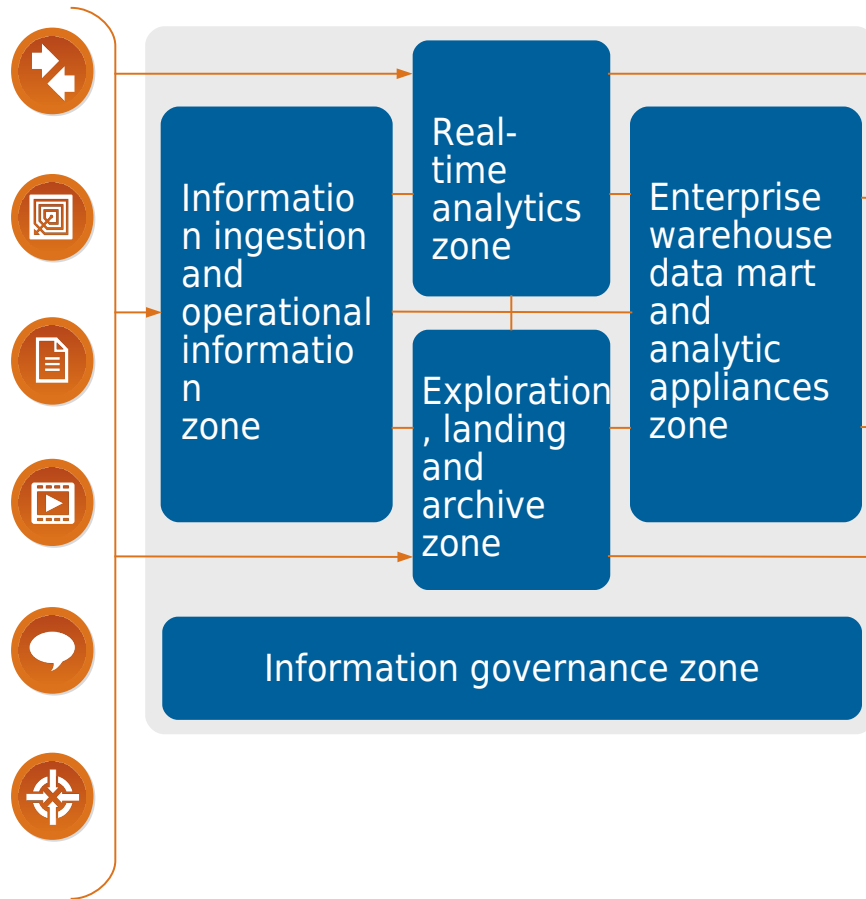


New/Enhanced Applications



# Realize It. A new architecture to leverage all data has emerged.

## All Data

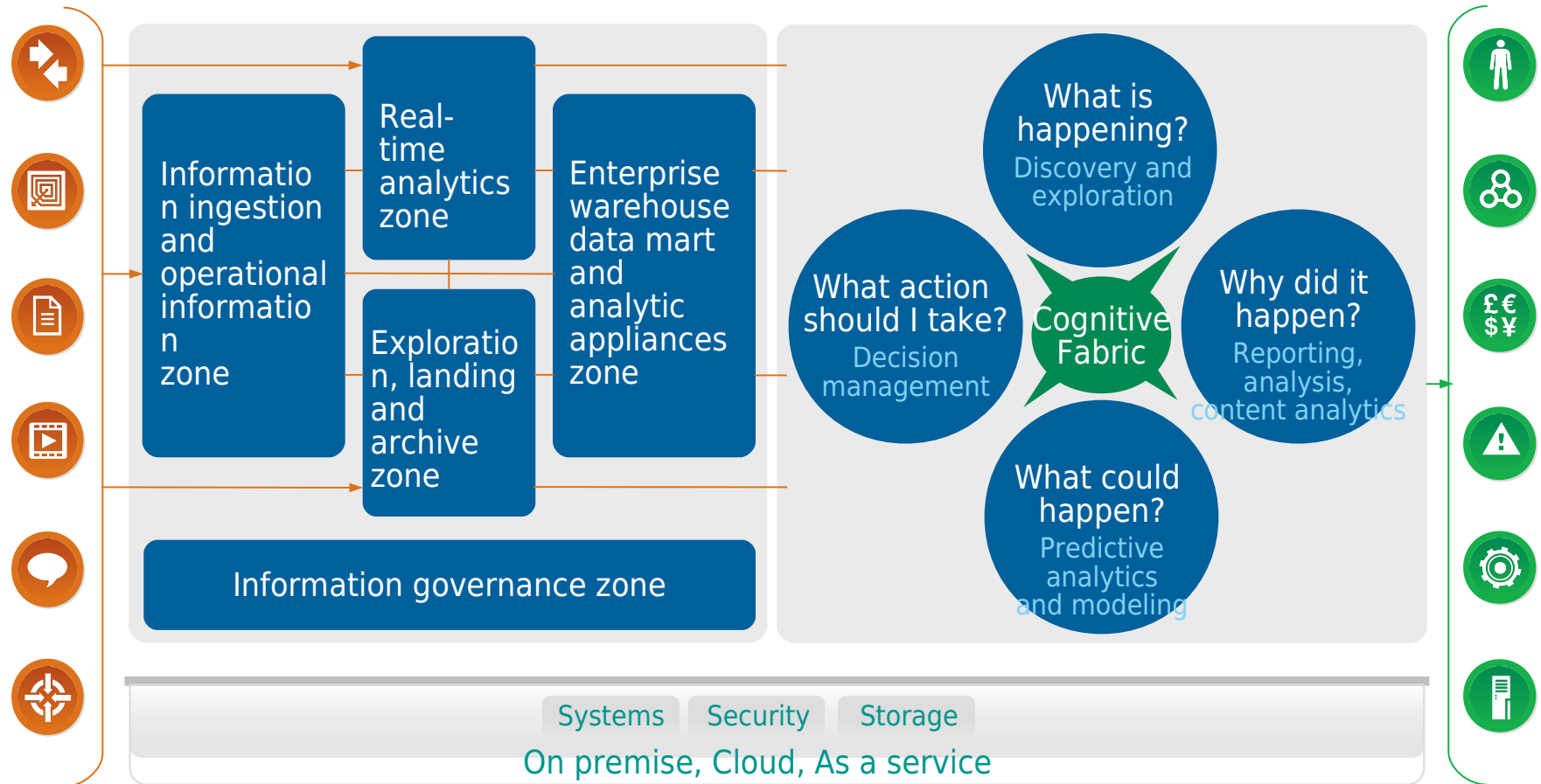


**Harness  
All Data  
& All  
Paradigms**

# Realize It. Implement a cognitive fabric powering all analytic applications

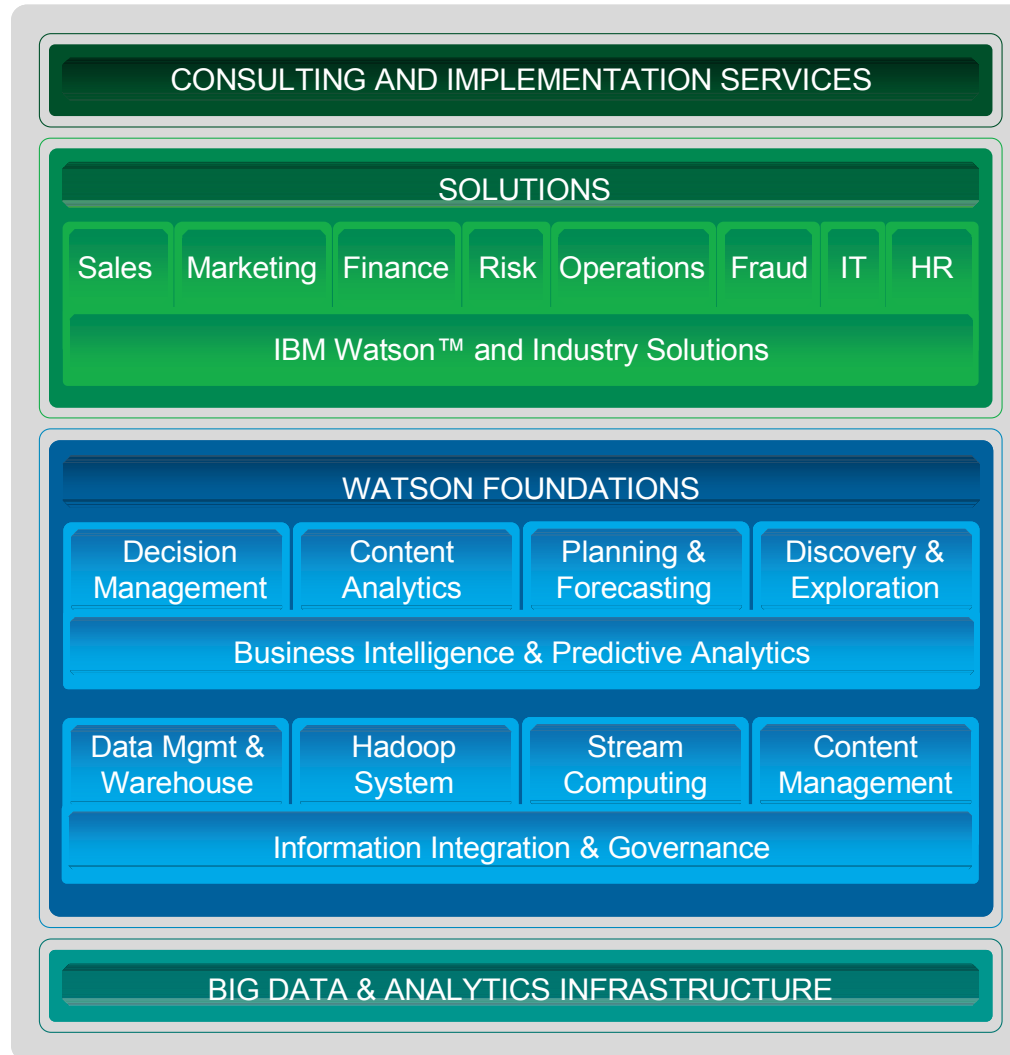
All Data

New/Enhanced Applications



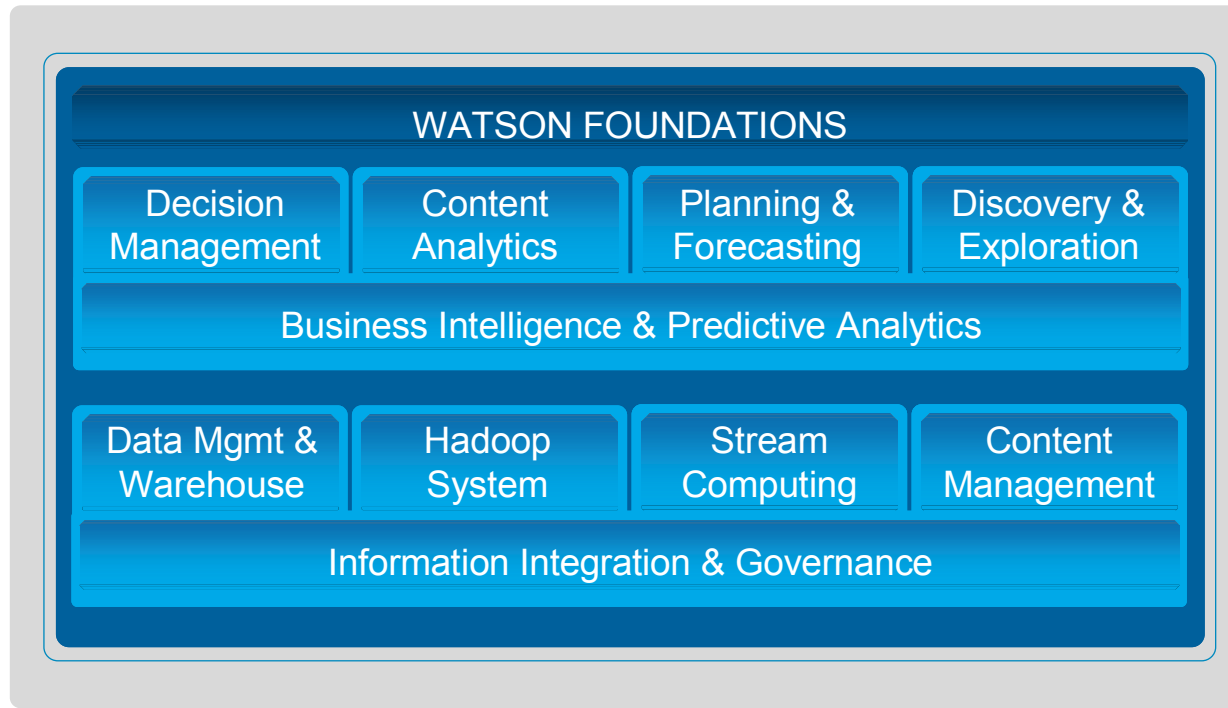
## Big Data & Analytics Infrastructure

# Realize It. Invest in best-of-breed big data & analytics platforms.



# Realize It. *Implement Watson Foundations.*

***IBM's new-era big data and advanced analytics platform, tool, and solution portfolio***



- ***Predictive and content analytics*** to uncover patterns not yet known
- ***Interactive exploration*** across all data
- ***Real-time analytics*** as data flows through an organization
- ***Best-of-breed big-data platforms*** optimized for performance, robustness, & manageability
- ***In-memory computing*** for speed of thought analytics
- ***Data governance*** across complete lifecycle
- ***Security and privacy*** with compliance



# Realize It. *The complete big-data ecosystem for modern business.*

## Watson-brand solutions

### Cross-industry

- Watson Engagement Advisor
- Watson Discovery Advisor
- Watson Analytics
- etc.

### Industry-specific

- Watson Health Cloud
- Watson Travel Cloud
- Watson Retail Cloud
- Etc.

### Ecosystem

- Watson Developer Cloud
- Watson Content Store
- Watson Talent Hub
- etc.

## Watson Foundations solutions

- IBM data warehousing solutions
- IBM Hadoop solutions
- IBM stream computing solutions
- IBM data management solutions
- IBM content management solutions
- IBM information integration & governance solutions
- IBM discovery & exploration solutions
- IBM BI & predictive analytics solutions
- IBM decision management solutions
- IBM content analytics solutions
- IBM planning & forecasting solutions

## Watson-enabled solutions

IBM and partner solutions that incorporate Watson technologies, solutions, and services

**Thank You**

## Legal Disclaimer

- © IBM Corporation 2014. All Rights Reserved.
- The information contained in this publication is provided for informational purposes only. While efforts were made to verify the completeness and accuracy of the information contained in this publication, it is provided AS IS without warranty of any kind, express or implied. In addition, this information is based on IBM's current product plans and strategy, which are subject to change by IBM without notice. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this publication or any other materials. Nothing contained in this publication is intended to, nor shall have the effect of, creating any warranties or representations from IBM or its suppliers or licensors, or altering the terms and conditions of the applicable license agreement governing the use of IBM software.
- References in this presentation to IBM products, programs, or services do not imply that they will be available in all countries in which IBM operates. Product release dates and/or capabilities referenced in this presentation may change at any time at IBM's sole discretion based on market opportunities or other factors, and are not intended to be a commitment to future product or feature availability in any way. Nothing contained in these materials is intended to, nor shall have the effect of, stating or implying that any activities undertaken by you will result in any specific sales, revenue growth or other results.
- If the text contains performance statistics or references to benchmarks, insert the following language; otherwise delete:  
Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.
- If the text includes any customer examples, please confirm we have prior written approval from such customer and insert the following language; otherwise delete:  
All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer.
- Please review text for proper trademark attribution of IBM products. At first use, each product name must be the full name and include appropriate trademark symbols (e.g., IBM Lotus® Sametime® Unyte™). Subsequent references can drop "IBM" but should include the proper branding (e.g., Lotus Sametime Gateway, or WebSphere Application Server). Please refer to <http://www.ibm.com/legal/copytrade.shtml> for guidance on which trademarks require the ® or ™ symbol. Do not use abbreviations for IBM product names in your presentation. All product names must be used as adjectives rather than nouns. Please list all of the trademarks that you use in your presentation as follows; delete any not included in your presentation. IBM, the IBM logo, Lotus, Lotus Notes, Notes, Domino, Quickr, Sametime, WebSphere, UC2, PartnerWorld and Lotusphere are trademarks of International Business Machines Corporation in the United States, other countries, or both. Unyte is a trademark of WebDialogs, Inc., in the United States, other countries, or both.
- If you reference Adobe® in the text, please mark the first use and include the following; otherwise delete:  
Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
- If you reference Java™ in the text, please mark the first use and include the following; otherwise delete:  
Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
- If you reference Microsoft® and/or Windows® in the text, please mark the first use and include the following, as applicable; otherwise delete:  
Microsoft and Windows are trademarks of Microsoft Corporation in the United States, other countries, or both.
- If you reference Intel® and/or any of the following Intel products in the text, please mark the first use and include those that you use as follows; otherwise delete:  
Intel, Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
- If you reference UNIX® in the text, please mark the first use and include the following; otherwise delete:  
UNIX is a registered trademark of The Open Group in the United States and other countries.
- If you reference Linux® in your presentation, please mark the first use and include the following; otherwise delete:  
Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both. Other company, product, or service names may be trademarks or service marks of others.
- If the text/graphics include screenshots, no actual IBM employee names may be used (even your own), if your screenshots include fictitious company names (e.g., Renovations, Zeta Bank, Acme) please update and insert the following; otherwise delete: All references to [insert fictitious company name] refer to a fictitious company and are used for illustration purposes only.