Data Symposium: Big Data

October 24th, 2014

Prepared for: Florida Department of Transportation (FDOT)

GARTNER CONSULTING

Gartner

Data Symposium: Big Data
Agenda

- What is Big Data?
  - Unpacking the lingo: Volume, Velocity, Variety
- Types of problems being solved using Big Data
  - Big data hype cycle review
  - Types of problems solved (across industries)
  - Types of big data analyzed
- Gartner experience with Big Data
  - Lessons Learned
  - Clearing up Myths and Realities
  - Gartner's Approach: Getting Started with Big Data

Gartner

What is Big Data?

We are more connected than ever – and this is generating HUGE amounts of data on a minute-by-minute basis

A picture of Facebook's network of relationships circa 2010 (based on 10MM samples)

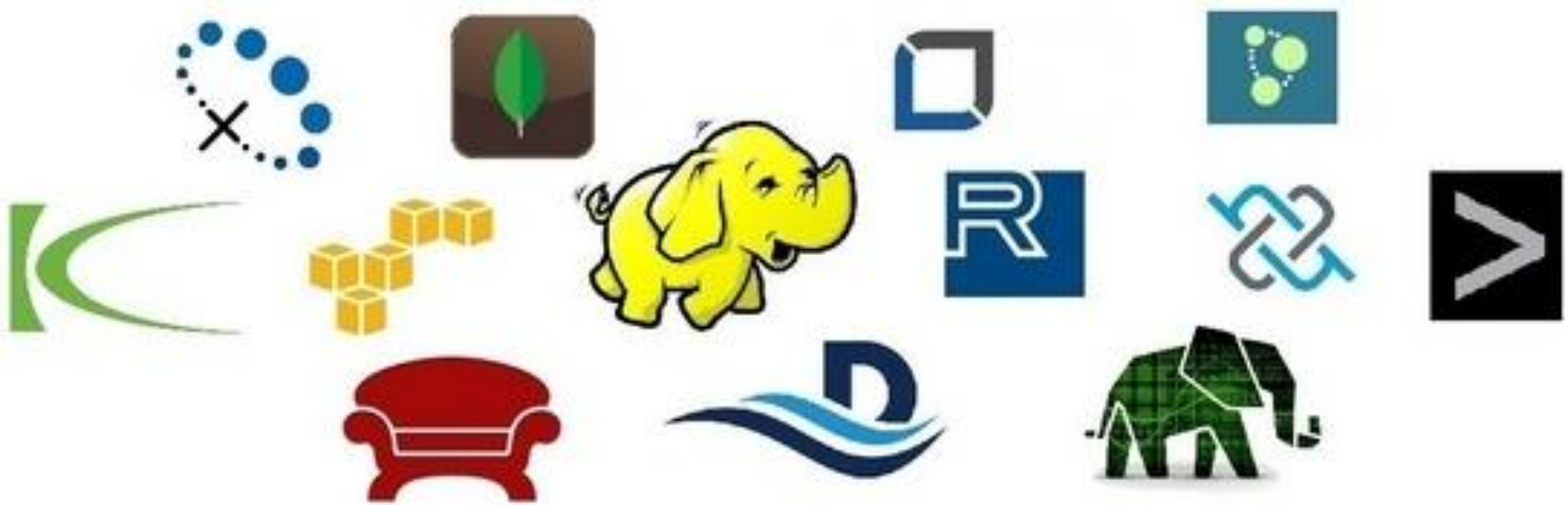
So, what does *Big Data* mean to you?

In your own words, what is Big Data?

*A bunch of **hype**?*

*A **problem** to be solved?*

*An **opportunity** to get closer to your customers and reinvent your business?*



*A **headache** to be managed?*

*A **question** waiting to be asked?*

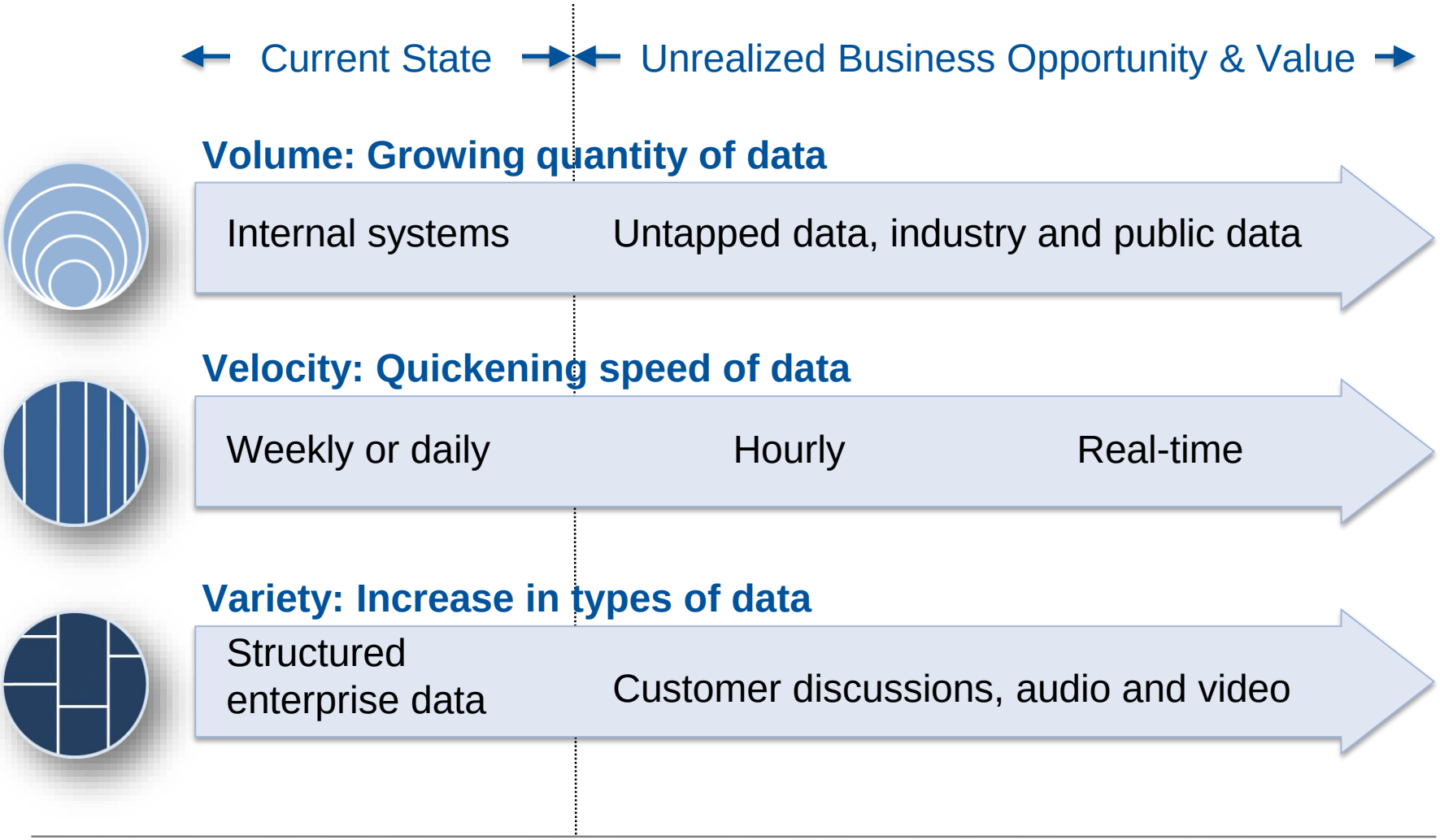
*A **technology** like Hadoop?*

Gartner defines big data as "high volume, velocity and/or variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation."

(Wow, that's a mouthful!)

(1) Unpacking the Lingo: “Volume, Velocity, & Variety…”

At it’s heart, Big Data = New Opportunities (for your customers & your business)



Gartner

(2) Unpacking the Lingo: “cost-effective, innovative forms of information processing…”

Storage and processing is cost-effective and scalable in ways that are significant

The Data, itself

- Historically, data was only static, **structured, and transactional** from enterprise systems
- Now, data is dynamic, **unstructured & contextual** which come from search, social media & mobile devices

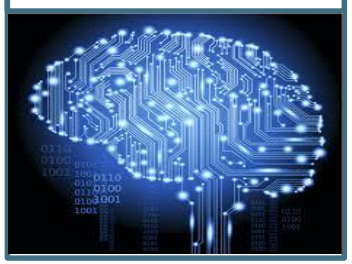
Visualization

- New forms of visualization techniques can better represent complex data like social network information
- New software vendors are using behavioral logic for an enhanced experience like calling attention with color

Storage

- **HDFS** (Hadoop Distributed File System) represents a big data storage framework that has marked improvements above **traditional relational DBs** as it leverages low cost servers and provides higher availability, scalability & fail-over tolerance.

Analysis



- Slower to lighting fast analysis
- New statistical programming languages are more robust
- **Data scientists** create machine-learning algorithms that automate work vs. the old days of data analysts taking days to crunch & report data

Processing

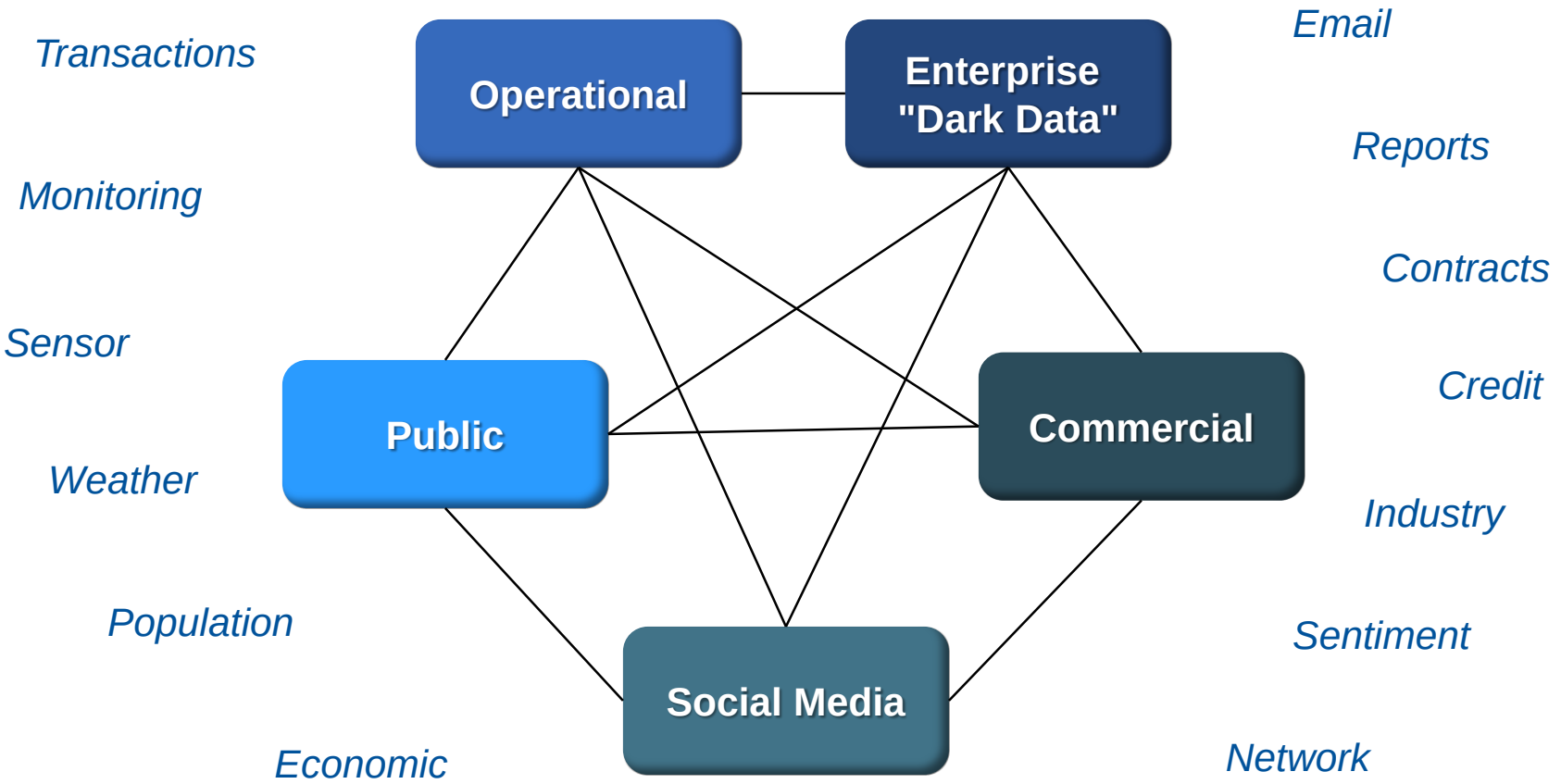
- **Hadoop MapReduce** offers an innovative framework that enables a machine’s compute power to be sent to where the data is housed in the cluster
- The Reduce step aggregates the data and resuts are sent back to the user w/ or w/o SQL.

Note: Big Data does *not* replace traditional data and analytics – it is an evolution of these practices, supported by both new technologies and an ever-increasing amount of data to analyze

Gartner

(3) Unpacking the Lingo: "…enables enhanced insight, decision making, and process automation"

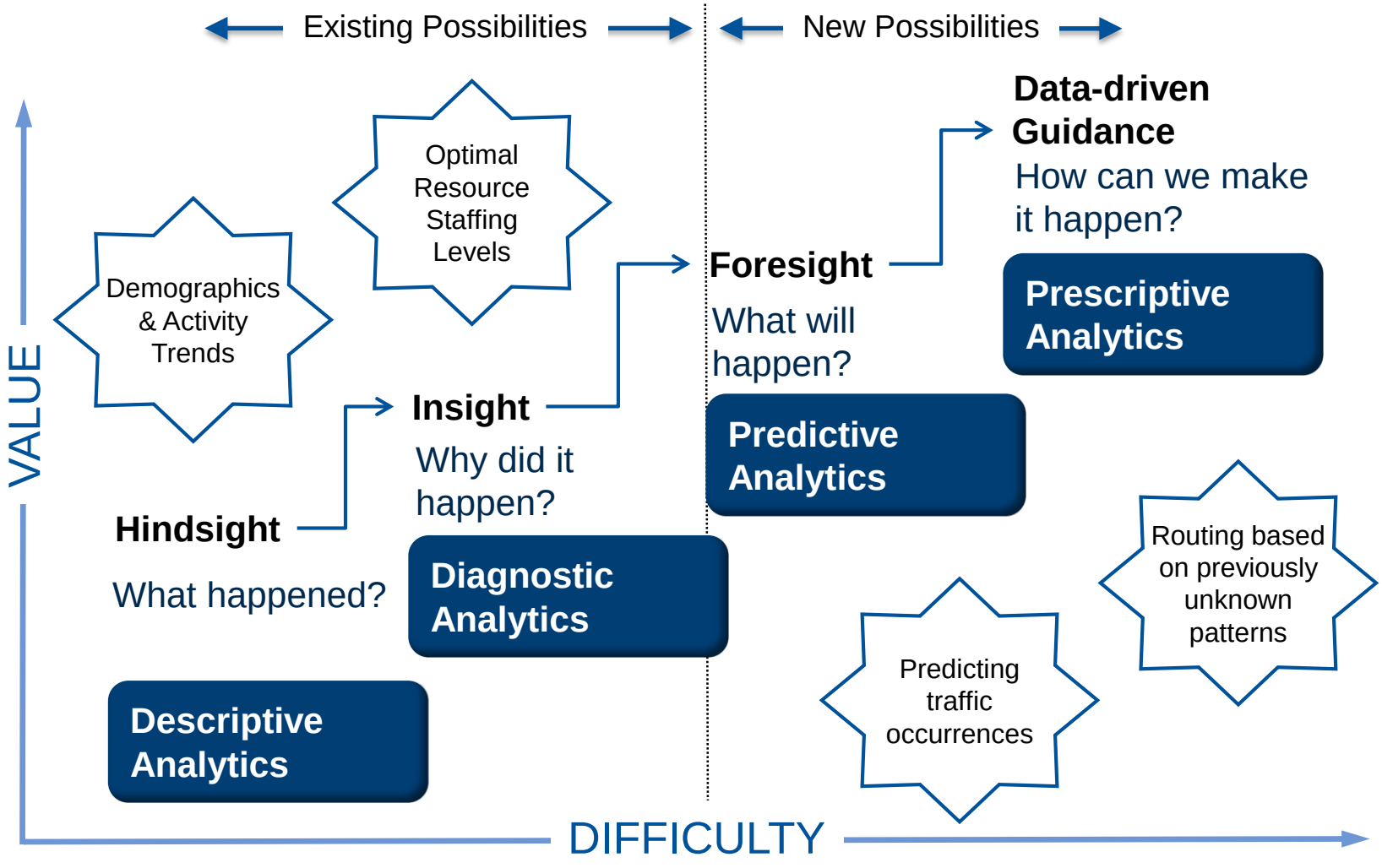
Big Data integrates, analyzes and correlates previously disparate data sources to unlock insights to "Big Questions"



Types of problems being solved with Big Data?

What types of questions can Big Data help answer?

Gartner Analytic Ascendancy Model



Potential Capabilities

Gartner

In fact, many leading companies are actively using Big Data to turn long-standing questions into actionable insight

Big Data Analytics in Action

Social Sentiment Analysis[1]

Recently purchased a company called Topsy in Dec 2013 for 200M. The company operates the world's only index of the public social Web in order track consumer commentary about brands, products and even predict events based on its unstructured big data.

Anticipatory Shipping[3]

In a patent filing, Amazon wishes to determine specific items to ship to individual customers before an order is placed by parsing structured and unstructured data.

Personalized Price Elasticity[2]

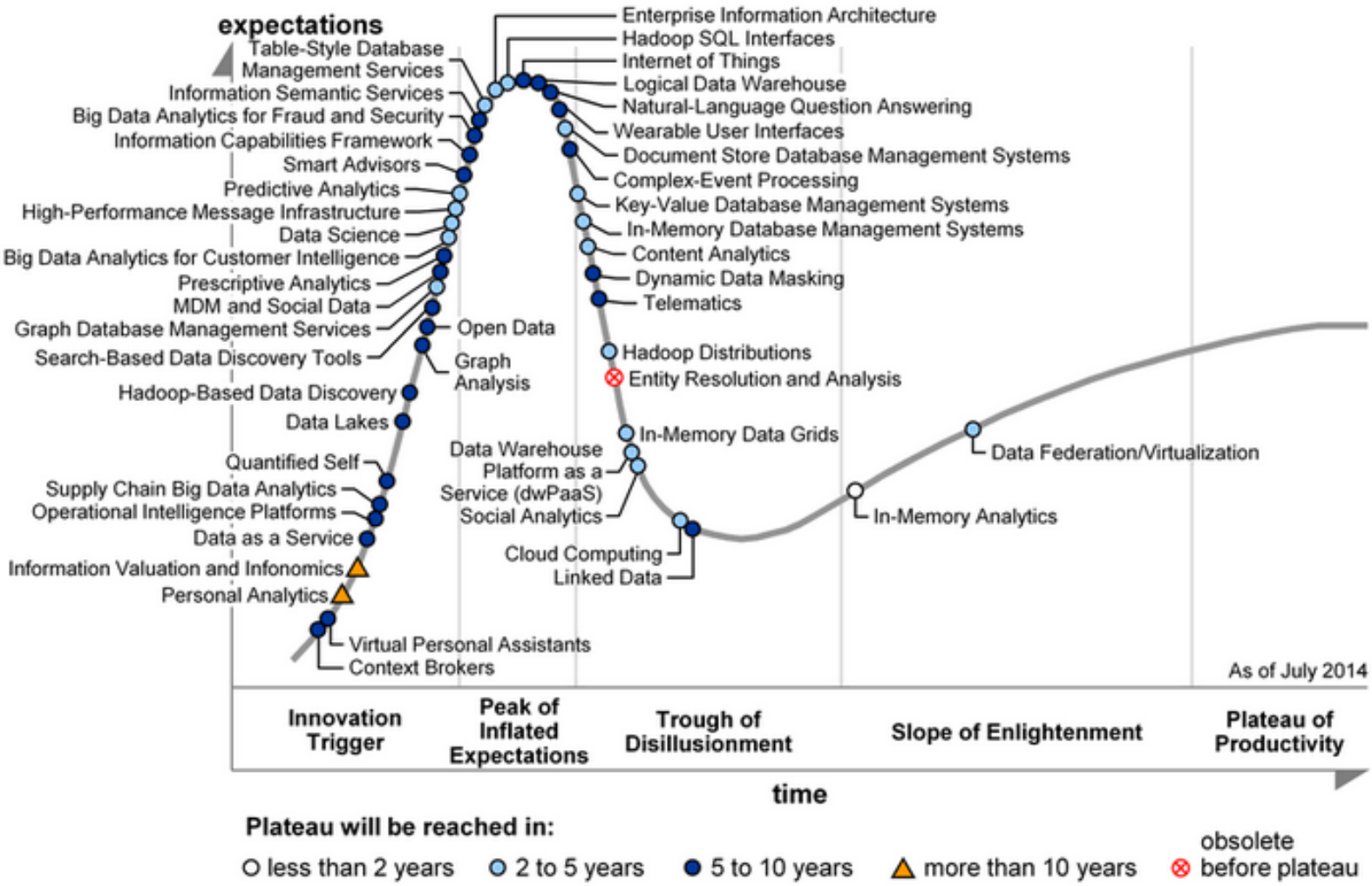
Is working on combining internet click stream and web log data with existing demographic information to run "willingness to pay" test so that it can tailor pricing based on individual rather than one price for all to increase revenue and market share.

Spending Habit Predictions[4]

After analyzing billions of records, it learned that when someone buys gas at around 4 p.m., the person will likely then spend $35-$50 on groceries or at a restaurant. The gas purchase didn't cause the food purchase.

1: Eweek; 2: Washington Post 3: USAToday 4: Business Insider

So – is the hype justified? Gartner's Big Data Hype Cycle, 2014



Yes! A few findings from Gartner's 2013 Big Data Study

- ***It's not just hype*** → *Big data investments in 2013 continued to rise*, with 64% of organizations investing or planning to invest in big data technology compared with 58% last year. Investments are led by media and communications, banking and services. Planned investments the next two years are highest for transportation, healthcare and insurance. Fewer than 8% of respondents have deployed.

- ***It's all about the customer*** → *Enhanced customer /citizen experience is the top big data priority*, with process efficiency close behind. Organizations struggle most with knowing how to get value from big data, compared with last year's top challenge of governance. Obtaining skills remains a critical issue for one-third of organizations.

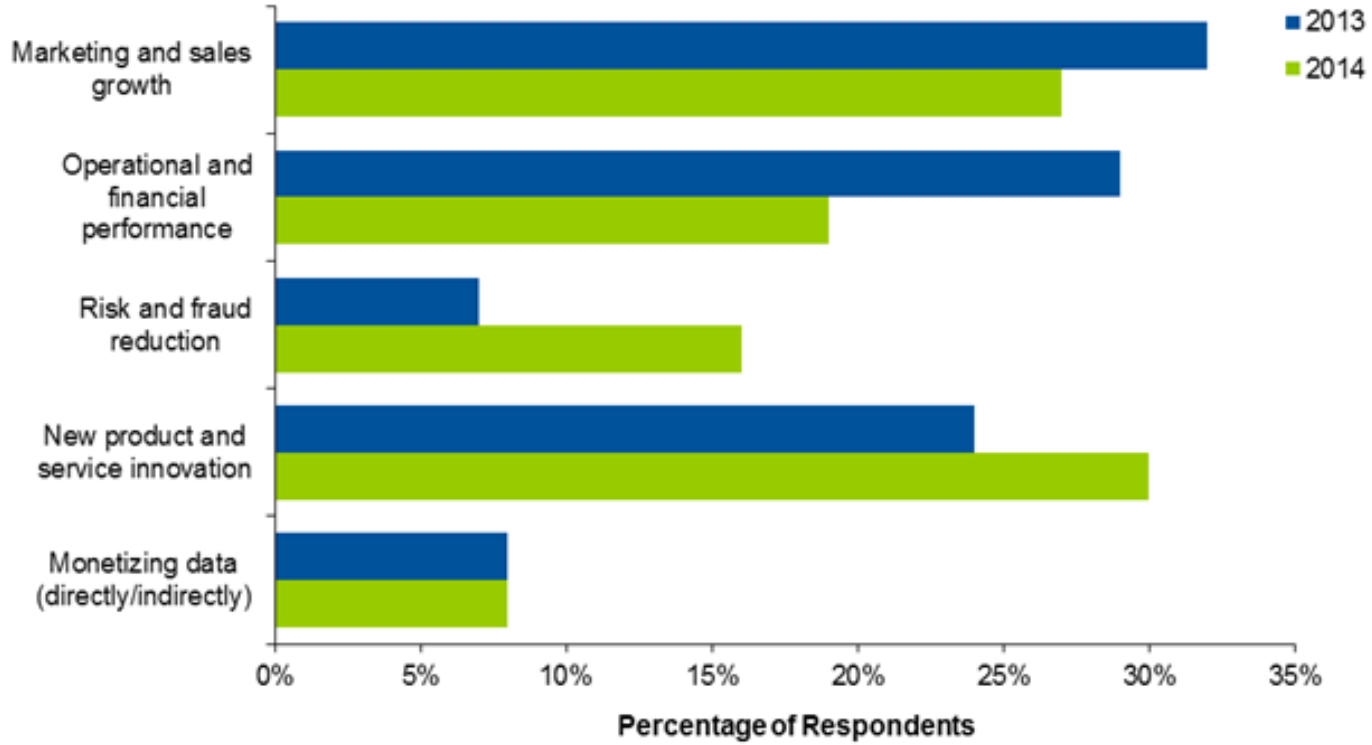
- **You already have all the data you need to start** → Big Data is touted as being about unconventional data sources and the use of new and innovative technologies; this is not yet reflected in the chosen sources for first projects. *Transaction and log data still dominates the big data being analyzed.*

- ***You don't have to throw out what you know*** → *Big data technologies supplement — but do not replace — existing information management and analytics*. As a result, cloud adoption, with its supplementary nature, is the overriding technology that companies are using to derive value from big data.

Gartner

So what is the biggest opportunity for Big Data?

Use of big data for direct customer-oriented activities such as sales and marketing remains high, big data is increasingly favored for new product and service development and indirect customer-oriented activities.



Data from Gartner's Big Data Webinars, March 2013 and 2014

Types of Big Data analyzed by industry

	Manufacturing and Natural Resources	Media/ Communications	Services	Government	Education	Retail	Banking	Insurance	Healthcare	Transportation	Utilities
Transactions	73%	62%	67%	67%	54%	93%	83%	81%	75%	79%	80%
Log data	44%	57%	58%	59%	54%	40%	66%	61%	33%	71%	60%
Machine or sensor data	53%	38%	35%	33%	31%	27%	27%	48%	42%	50%	40%
Emails /documents	27%	43%	43%	41%	46%	27%	34%	39%	17%	29%	20%
Social media data	32%	52%	39%	26%	54%	73%	27%	13%	-	50%	-
Free-form text	17%	24%	28%	30%	31%	20%	34%	35%	67%	21%	40%
Geospatial data	27%	14%	19%	19%	38%	27%	27%	26%	8%	29%	40%
Images	19%	24%	17%	11%	38%	13%	5%	16%	25%	7%	-
Video	8%	29%	12%	7%	31%	13%	-	6%	8%	7%	-
Audio	10%	19%	8%	4%	8%	-	-	6%	-	-	-
Other	8%	14%	13%	15%	8%	7%	10%	16%	42%	14%	-
n =	*59*	*21**	*127*	*27**	*13**	*15**	*41*	*31*	*12**	*14**	*5**

Gartner

Gartner experience with Big Data

Gartner

Lessons Learned based on typical failures

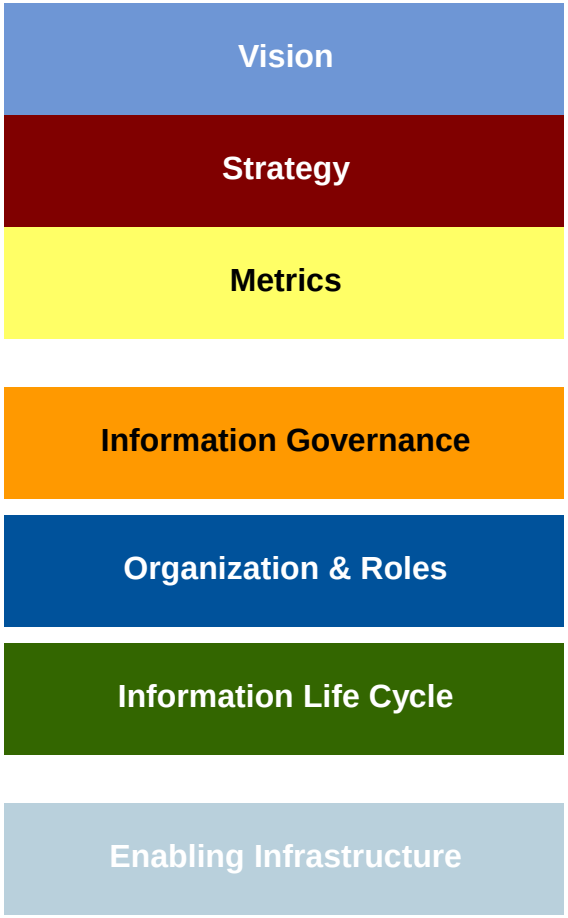
Failure Node		Impact	Gartner Approach
Premature focus on the solution	Many organizations leap directly to engineering a solution.	Failure to achieve success due to absence of business value and sustainability.	Big Data/ BI has a single purpose: achieve business goals or address business deficiencies. The departure point for any EIM program are the following: business strategy, business unit plans, business drivers, IT drivers, and business painpoints. It is the cumulative effect of all of these factors that leads to the formulation of a sustainable solution.
Myopic focus on tools	Many organizations consider EIM to be a purely technical problem that can be addressed through the purely technical solution.	Failure to achieve success due to tool limitations and poor implementation resulting from lack of business participation.	Big Data/ BI is a discipline the entails seven interrelated blocks: vision, strategy, metrics, governance, organization, processes, and technology infrastructure. These have to be developed in this specific order with technology following all the rest.
Disregard of the culture	Many organizations fail to accommodate the cultural dimension.	Failure to achieve success due to unimplementable structures, ineffective governance, and adoption resistance.	Big Data/ BI has to always follow the cultural norms and organizational particularities of the enterprise. It has to adequately balance a large number of conflicting factors: responsiveness v. control, centralization v. decentralization, strategic goals v. immediate needs, enterprise value v. local utility.
Excessive scope	Many organizations fail to control their scope and arbitrarily elect to encompass everything under their EIM program.	Failure to achieve success due to increased complexity and lack of opportunity to incrementally increase maturity.	Big Data/ BI scope is determined as a series of progressive priority curves that segment the ease of implementation v. affinity plot. This ensures that the enterprise focus in narrow in scope and incremental in evolution that add business value while allowing the enterprise to learn.
Lack of change management	Many organizations fail to consider change management or they discontinue it after the initial deployment.	Failure to achieve success due to lack of business buy-in or misuse of EIM practices.	Change management is the very first activity that an Big Data/ BI program has to accomplish. It entails marketing, training, and incentives and its purpose is to effect change at the enterprise, project, and individual level.

Clearing up Myths and Realities

Common myth	Reality
Big data = Hadoop	Big data is a technology stack that satisfies the definition of big data. It could include Hadoop but is not limited to it. Big data technologies include, but are not limited to, streaming data, search, NoSQL and RDMS.
Everybody is implementing big data technologies and we're behind!	The majority of enterprises, including those touted by the media as successful adopters of big data technologies, are at Stage 2 (Aware) and Stage 3 (Opportunistic).
A single data scientist will implement a big data solution for our enterprise.	A multidisciplinary team with diverse skills can meet technology challenges and solve complex business problems of big data adoption.
Companies can completely migrate off RDBMSs.	Don't forfeit RDBMS technologies. They serve many purposes better than big data technologies. Big data offerings remain immature compared to RDBMSs; RDBMSs are an important component of the LDW.
The ratio between unstructured and structured data is 80:20. Therefore, unstructured data contains 80% of information value.	80% represents the share of the quantity of unstructured data, not necessarily the value of it. Structured data has been refined, and its density and quality are much greater than comparable amounts of unstructured data.

Gartner

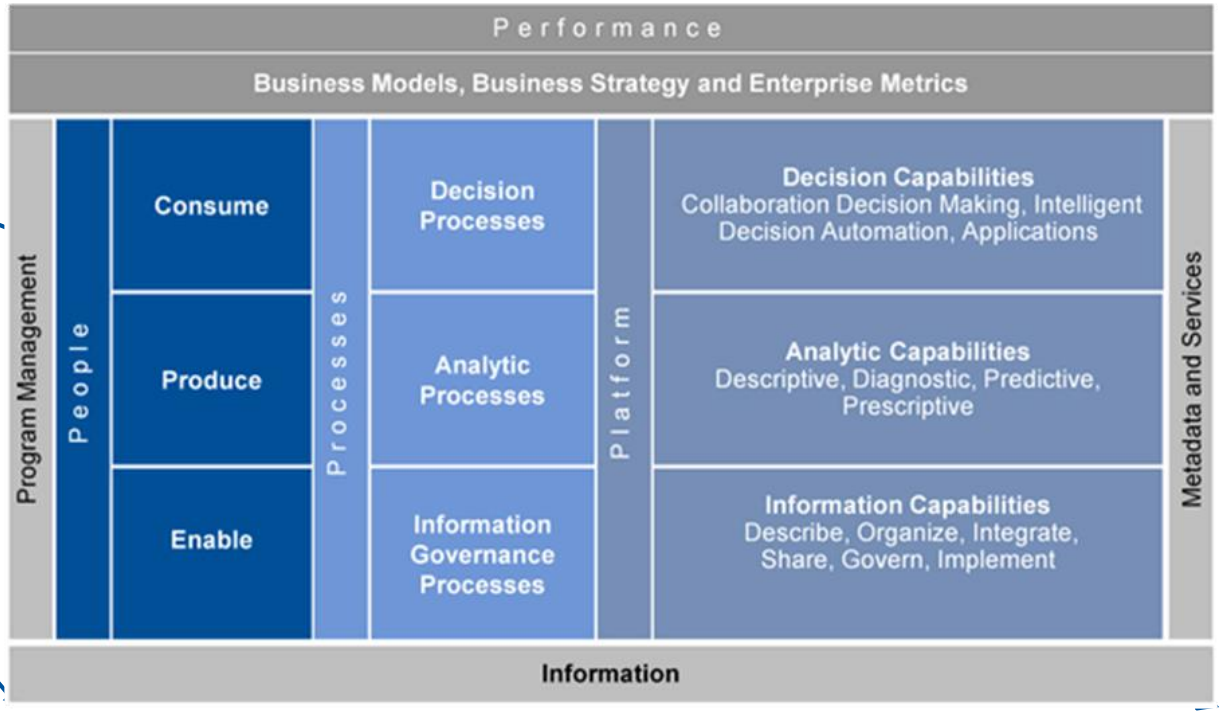
Gartner's Approach to Getting Started
Gartner's Seven Building Blocks of BI/Big Data



- Gartner has established a formal framework to guide technology choices with equally important organizational, governance, and process dimensions. The elements of this framework reinforce the belief that effective Business Intelligence capabilities do not depend on technology alone. In fact, BI/Big Data capabilities should be seen as an integral part of a larger effort to improve performance at all levels of an organization.
- The framework comprises the following components:
  - Vision – Describes what BI looks like, why it is needed, and how it supports the business vision
  - Strategy – Articulates how the BI vision will be realized
  - Metrics – Captures how the progress and business contributions of BI are measured
  - Information Governance – Establishes the decision rights framework and assigns roles and responsibilities
  - Organization & Roles – Forms the groups and organizational structures to support BI
  - Information Life Cycle – Sets up the processes to support BI
  - Enabling Infrastructure – Provides the technology capabilities to meet the BI needs

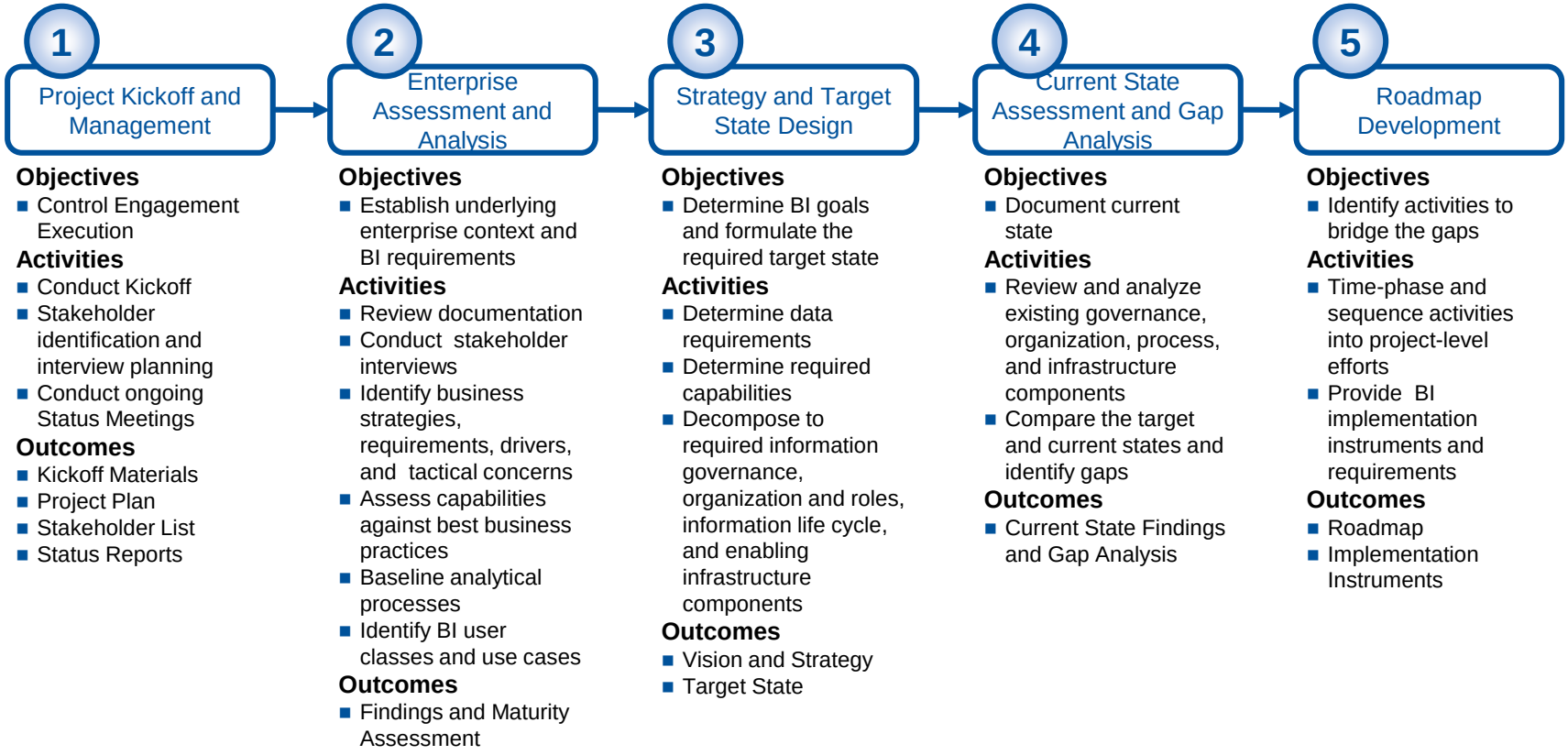
Gartner's Business Analytics Framework

- The Gartner's Business Analytics Framework provides the detailed dimensions of BI and DW and maps their particularities to the BI building blocks.
  - Provides the dimensions for the maturity assessment
  - Establishes the reference for the target and current state descriptions and the any gap analysis
  - Provides the interrelated components that evolve the BI/DW governance model and framework
  - Frames the interrelationship among BI and other analytical capabilities such as Corporate Performance Management and establishes a unifying framework



Gartner has the right approach and strong experience to support the successful delivery of a BI/Big Data Strategy and Roadmap

- Gartner understands how to approach BI/Big Data assessments. The figure below represents a "standard" Strategy and Roadmap Formulation approach.



Thank You!

Paul Jurevicius
Director , Business Intelligence &
Enterprise Info. Mgmt.
Gartner Consulting
Telephone: +1 312.508.2264
Paul.jurevicius@gartner.com

Gartner