June 17, 2015

Innovate Forward

Data Isn't Everything

The Challenges of Big Data, Advanced Analytics, and Advance Computation Devices for Transportation Agencies.
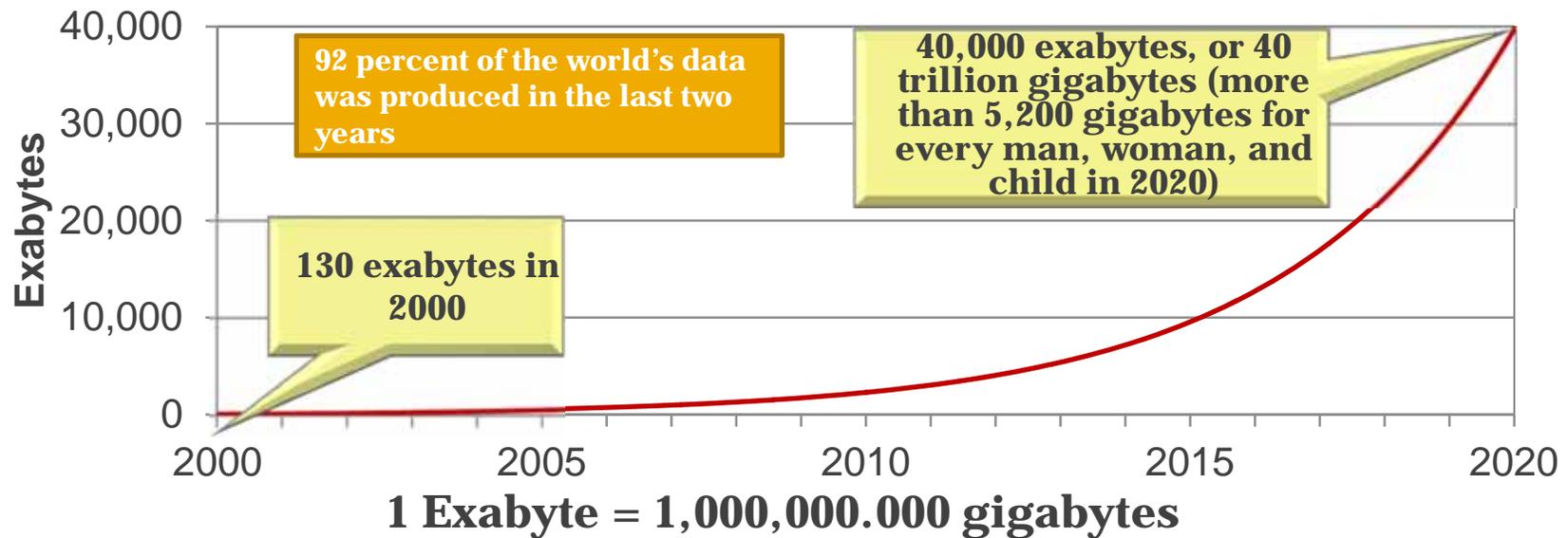Using Data to Support Mission, Administration, and Operations

Booz | Allen | Hamilton

Purpose

+ **This presentation explores the challenge of Big Data and advanced analytics for transportation agencies**

+ **Specifically it:**

  – Provide background to the Big Data and advanced analytics revolution

  – Identifies specific challenges

  – Discuss approaches to addressing these challenges

Background: The Big Data Revolution

Over the next decade we will see an explosion in data from omnipresent sensors (>50 billion passive and active sensors by 2020) and other data collection technologies. Combined with a vast increase in processing power and a continuing decline in cost, this offers the possibility for dramatic growth

92 percent of the world's data was produced in the last two years

40,000 exabytes, or 40 trillion gigabytes (more than 5,200 gigabytes for every man, woman, and child in 2020)

130 exabytes in 2000

1 Exabyte = 1,000,000.000 gigabytes

Background: Will the Big Data and advanced analytics live up to its promise?
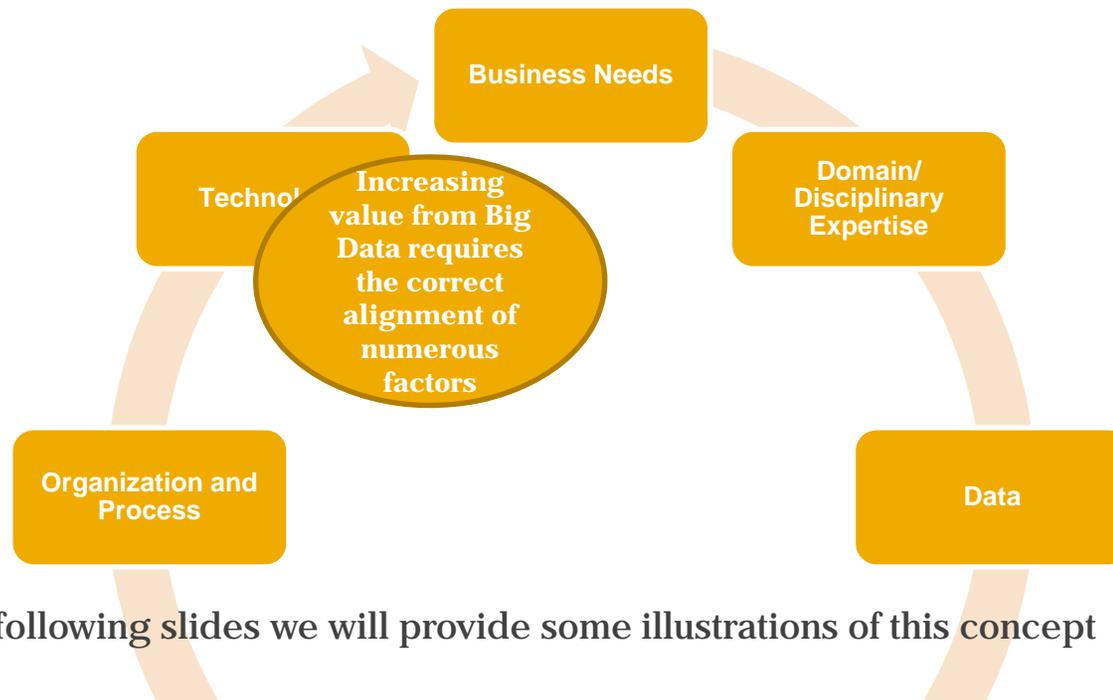
The explosion of Big Data and processing power has generated a lot of hype and offers the potential that advanced analytics techniques could be easily accessible to everyone- generating substantial value

The key questions are how should these technologies be used to maximize their value and what are the key factors (technological, organizational, cultural) that lead to organizations being able to realize their promise

The Challenge of Big Data and Advanced Analytics:

We have found that organizations gain the most from advanced analytics and Big Data when they consider a number of factors
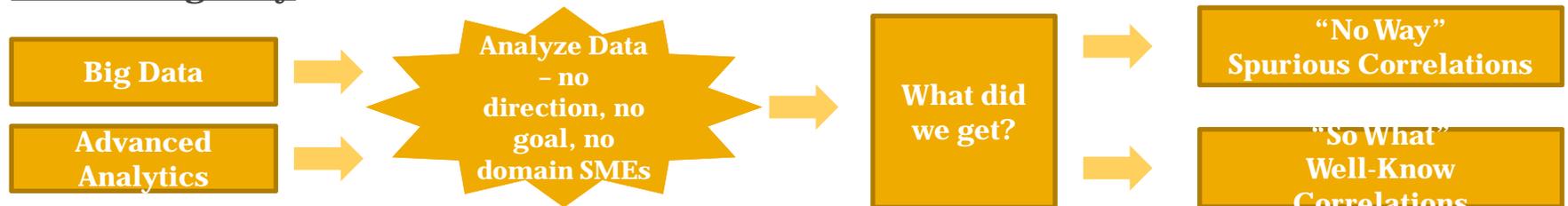
Business Needs

Domain/
Disciplinary
Expertise

Technol...

Increasing value from Big Data requires the correct alignment of numerous factors

Data

Organization and Process

On the following slides we will provide some illustrations of this concept
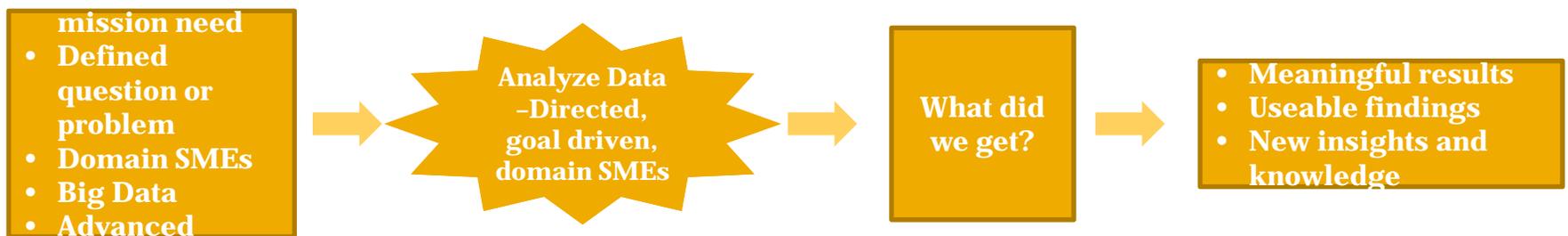
4

Example: There's No Substitute for Domain SMEs and a Carefully Defined Question

Too often organizations implement Big Data and advanced analytics initiatives without integrating existing domain expertise into the initiatives or defining the key questions or mission/business needs they need to have answered
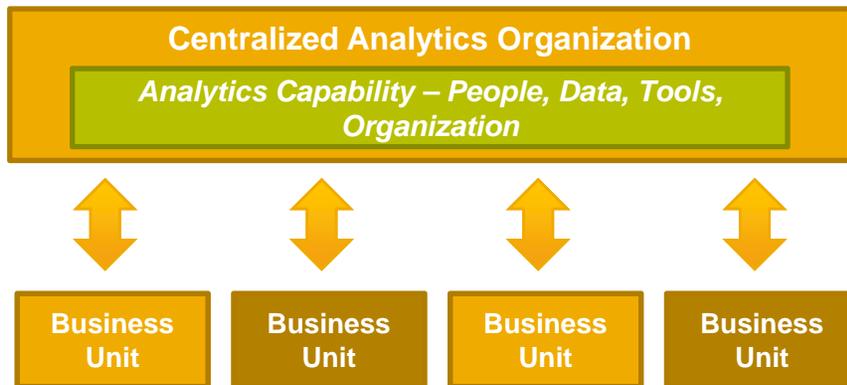
The "Wrong" Way:

| Big Data | → | Analyze Data – no direction, no goal, no domain SMEs | → | What did we get? | → | "No Way" Spurious Correlations |
| Advanced Analytics | → | | | | → | "So What" Well-Know Correlations |

The "Right" Way:

- mission need
- Defined question or problem
- Domain SMEs
- Big Data
- Advanced

→ Analyze Data –Directed, goal driven, domain SMEs → What did we get? →

- Meaningful results
- Useable findings
- New insights and knowledge

Example: How can you organize analytics?

There are many ways to organize analytics – each has different advantages and disadvantages

CENTRALIZED ANALYTICS MODEL

Centralized Analytics Organization

Analytics Capability – People, Data, Tools, Organization

Business Unit | Business Unit | Business Unit | Business Unit

- Critical mass of resources to catalyze performance
- Access to cross-functional data to conduct high-value analytics
- Avoid duplication
- Reduces costs and increases analytics output
- Organizational resistance
- Needs to work with domain experts
- Need to resolve data ownership/stewardship

BUSINESS UNIT ANALYTICS MODEL

Business Unit | Business Unit | Business Unit | Business Unit

Analytics Capability (within each Business Unit)
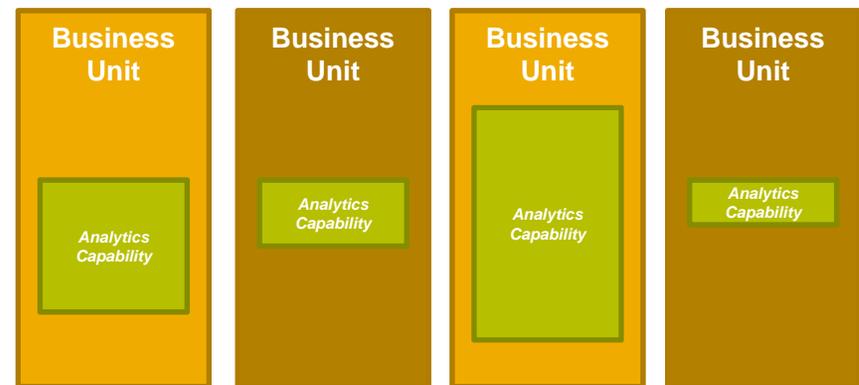
- Insufficient resources in each business unit to catalyze performance
- Lack cross-functional data to conduct high-value analytics
- Duplication of capacity – competition for scare resources
- Increases costs and lowers analytics output
- Domain expertise and analytics combined in single organization
- Data ownership/stewardship is simplified

Example: How can you organize analytics?

	Description	Pros	Cons
Centralized	Centralized analytics group providing analytical services to the organization	• Critical mass of resources to catalyze performance • Access to cross-functional data to conduct high-value analytics • Avoid duplication • Reduces costs and increases analytics output	• Organizational resistance from program offices • Needs to develop means to work with domain SMEs • Need to resolve data ownership/ stewardship • Requires charge-back mechanism
Center of Excellence	Analysts are distributed through the organization but are supported by a small group responsible for training, technology tracking, disseminating best practices, and facilitating communication	• Analysts and domain SMEs develop natural, organic cooperative relationship • Allows organization to share best practices and core analytics function • Consumer driven – minimal organizational resistance • Data ownership/stewardship is simplified	• Lack cross-functional data to conduct high-value analytics • Duplication of capacity – competition for scare resources
Decentralized	Analysts are dispersed among the organization and aligned with business units – not centralized control or coordination	• Analysts and domain SMEs develop natural, organic cooperative relationship • Data ownership/stewardship is simplified • Each organization get the analytics it perceived it needs	• Insufficient resources in each program office catalyze performance • Lack cross-functional data/platform to conduct high-value analytics • Duplication of capacity – competition for scare resources

Example: Data Governance

Proper Analytics governance can prevent major problems from emerging. For example:

Problem:
"Cowboy Analytics"

Models and techniques are used by users who do not understand their purpose, requirements or the limitation of their data. The result is that resources may be wasted, quality stuffers and the reputation of analytics organizations can be damaged when substandard products are produced
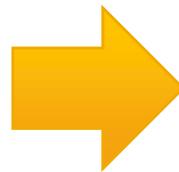
Solution:
"With Great Power Comes Great Responsibility"

Governance needs to establish training requirements, recommended technology tools, verification and validation protocols, and risk management processes to keep enterprise-wide analytics refreshed and reliable

Example: Big Data vs. Smart Data

Data and advanced analytics are meaningless unless business needs and domain expertise provide context and identify problems of interest. Smart Data mean something to domain SMEs, where key messages can be communicated to non-SMEs through data visualization and messaging

Data Transformation

DUMB DATA

- Difficult to acquire
- Hard to integrate
- Numerous errors
- Difficult to understand

- **Business needs define the key questions to answer (data don't tell you what you need to ask)**
- **Domain expertise adds context and meaning**
- **Data cleansing and preliminary analytics create usable data**
- **Intelligent integration creates meaningful cross functional data sets**

SMART DATA

- Contextual
- Easy to access
- Integrated
- Accurate
- Easy for SMEs to understand

Example: Don't Use a Wheel to Break a Butterfly

Technology is making numerous advanced analytical techniques available, but they need to be combined with the appropriate data, domain expertise and skilled users – not all problems require complex solutions
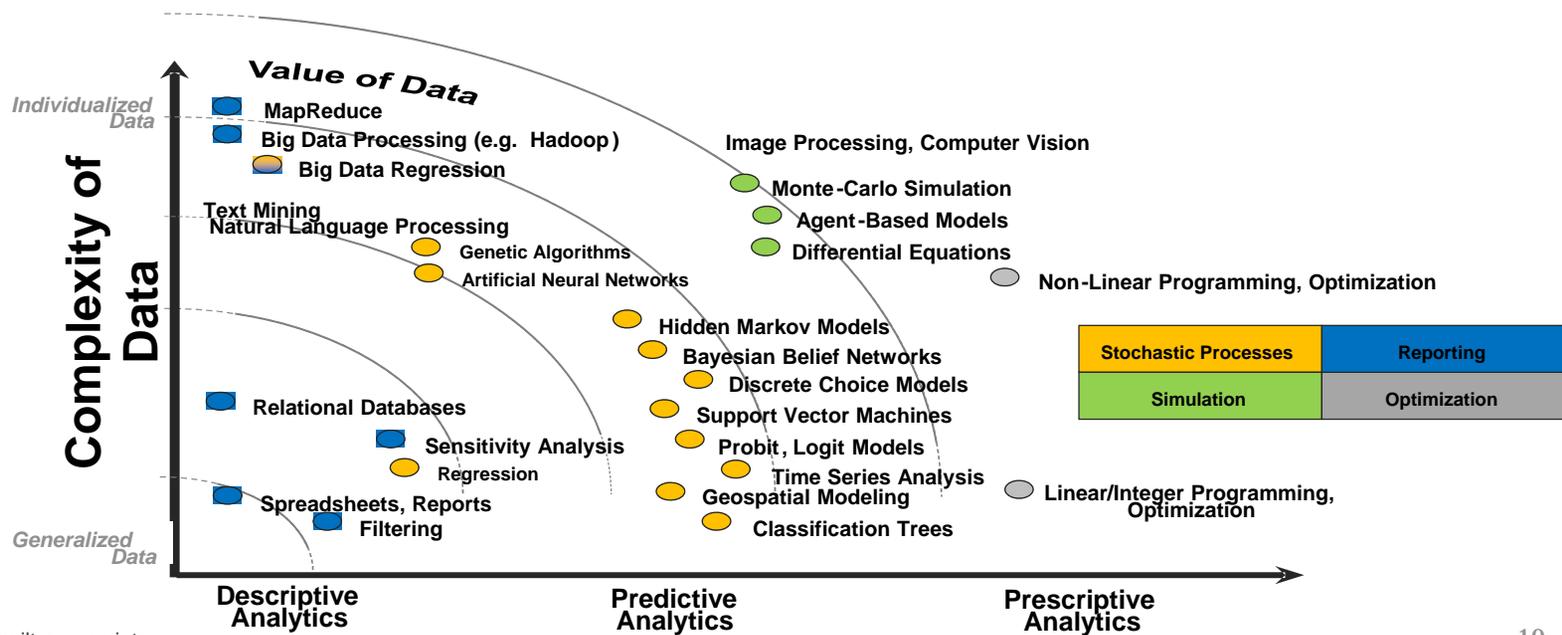
Key takeaways from our experience implementing Big Data and advanced analytics initiatives

Context is King: Big Data and advanced analytics are nothing without domain expertise

No Substitute for a Passionate Question: Business and mission needs and a desire to understand a critical phenomena determine value – data doesn't speak for itself

Organization and Culture Matter: How advanced analytics is integrated into an organization matters –analyze and plan your adoption of new technologies

Data Ownership/Stewardship is Critical: Who owns the data, the access rights that analysts have and the responsibility for managing data are critical

Merging Expertise: Merging domain and analytics expertise is difficult – finding the right people and the organizational processes that work are often the biggest problem

Don't Pay for What you Don't Need: Big Data and advanced analytics are not for every organization or every problem – match the tool to the challenge

Advanced Analytics & Data Dashboard (A^2D^2) Can Help Organizations Understand Advanced Analytics

+ Booz Allen has developed the Advanced Analytics & Data Dashboard (A^2D^2) approach to assess the readiness of an organization to implement advanced analytics and to help organization leaders decide the path forward in developing advanced analytics

+ It identifies the gap between current capabilities and goals that are developed from business needs by:

  – Identifying technology and data requirements to support business needs

  – Assessing organizational and cultural issues that impact on an organization's ability to support advanced analytics

+ A^2D^2 provides a rapid, stakeholder-focused diagnostic that gives organization map of how to move from their current "as-is" state to a envisioned "to-be" state where Big Data and advanced analytics can be used to support business needs
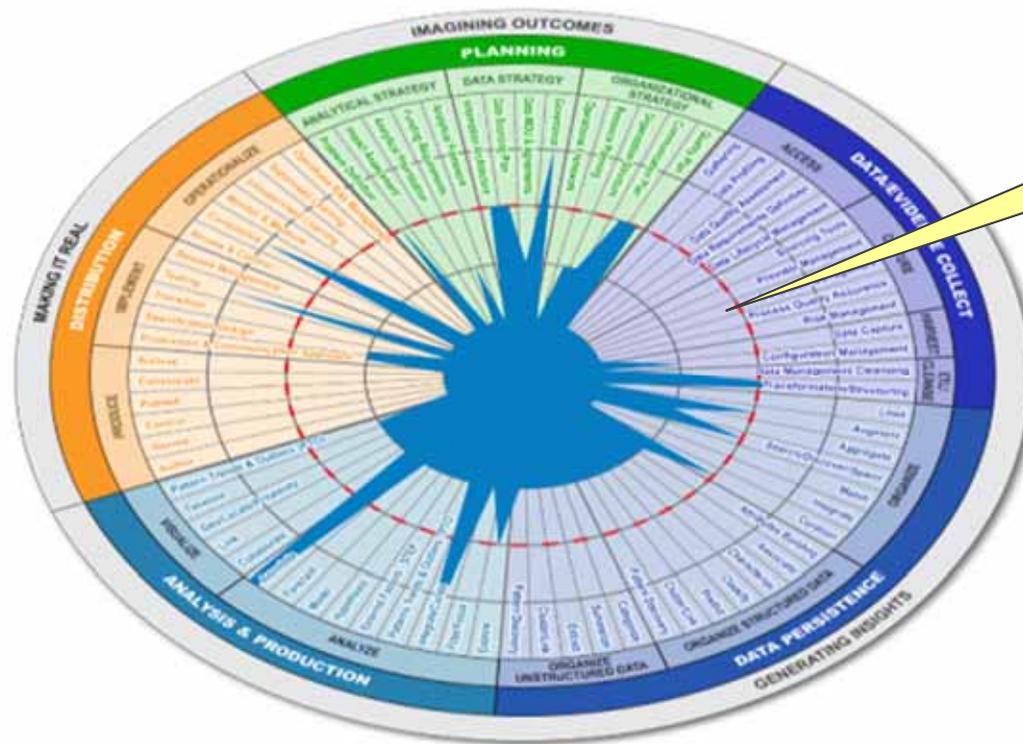
Advanced Analytics & Data Dashboard (A²D²) Framework

- A²D²'s **Analytical Framework** is the "Lens" through which all activities are evaluated

- The "right" framework is the one that facilitates communication and understanding across all stakeholders.

- The final model captures the *iterative*, *highly fragmented*, and *personal* nature of analytical activity and identifies key types of analyses:

  - Exploratory
  - Descriptive
  - Predictive
  - Decision
  - Operational
  - Deductive
  - Inductive
  - Abductive

- The operational perspective of the analytical activity is defined in the **Operational Framework**

  - Exposes organizational boundaries and structures and their impact on analysis

  - Technology and associated products can be mapped onto Operational Model

  - Capability gaps are identified

  - Identifies capabilities that enable the IT platforms to interact with operational components to support analytical tasks

PLANNING

Plan & Manage
- Manage Analysis
- Manage Collection
- Control Quality

DATA/EVIDENCE COLLECT
- Access
- Capture
- Harvest
- ETL/Cleanses
- Code: Geo, Temporal, etc.
- Translate

Organize
- Load/Persist
- Aggregate
- Match
- Search & Select
- Attach Attributes

DISTRIBUTION

Produce
- Author
- Control
- Review
- Publish
- Collaborate (Drill Down/Drill Back)
- Archive

Collaboration and Knowledge Management

Visualize
- Link
- Geo/Locate/Proximity
- Timeline
- Trends
- Anomalies

Analyze
- Explore
- Filter/Focus
- Hypothesize
- Model
- Forecast
- Annotate/Enhance
- Peer Review/Validation

Structured Data
- Associate
- Characterize
- Classify
- Predict
- Cluster
- Discover Patterns
- Analyze Execution

Unstructured Data
- Categorize
- Cluster
- Extract
- Summarize

ANALYSIS & PRODUCTION

DATA PERSISTENCE

13

A²D² illustrative work product: Demonstrates multiple dimensions of competency

Summary Chart indicating current capabilities and areas of focus (in **Blue**), and the target capability level denoted by the red dashed circle

15