November 3rd, 2016

Innovate Forward

HARNESSING THE POWER OF YOUR DATA

Jesus Jackson – Chief Data Scientist

Booz Allen Hamilton and Client proprietary and business confidential

Booz | Allen | Hamilton

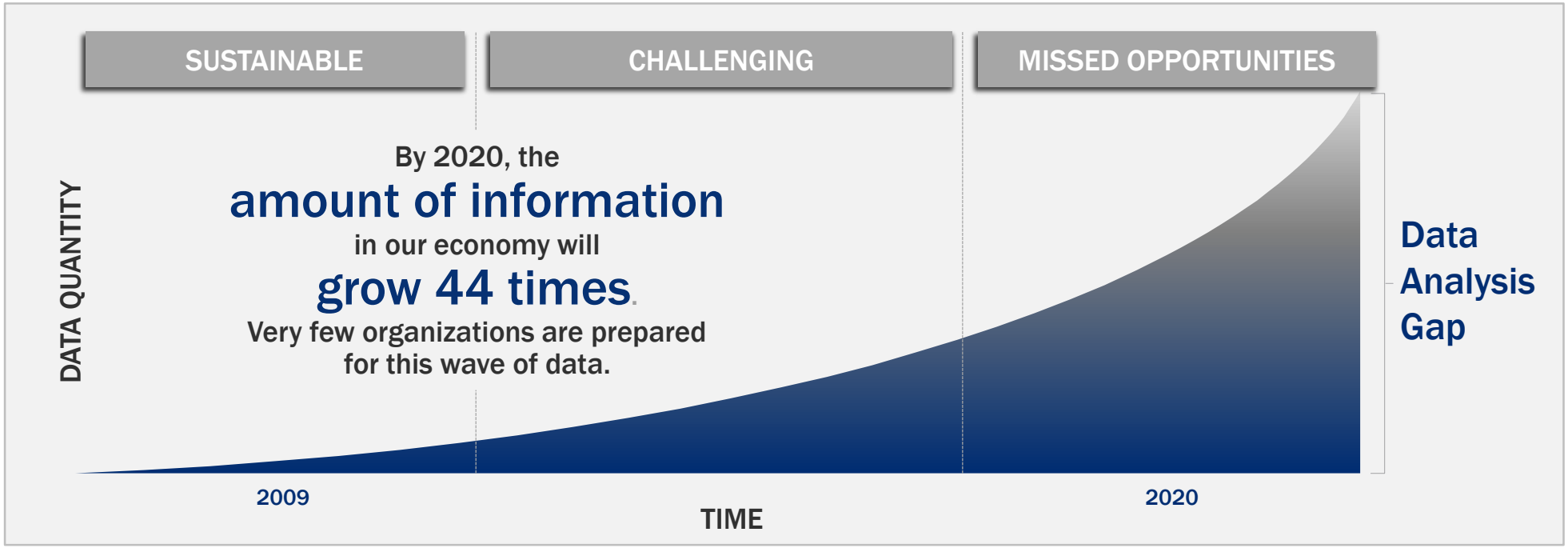
Who Am I?

- I've been with Booz Allen for 7.5 years and I help drive our data science practice
- I've had a deeply technical career and led projects that covered broad technical areas. My focus the past few years has been on large-scale big data and advanced analytics platforms (think Hadoop, Spark, etc).
- Fun fact: I play drums and have been playing for over 12 years. My neighbors love me :)
- All right…let's talk about data!

Booz Allen Hamilton and Client proprietary and business confidential

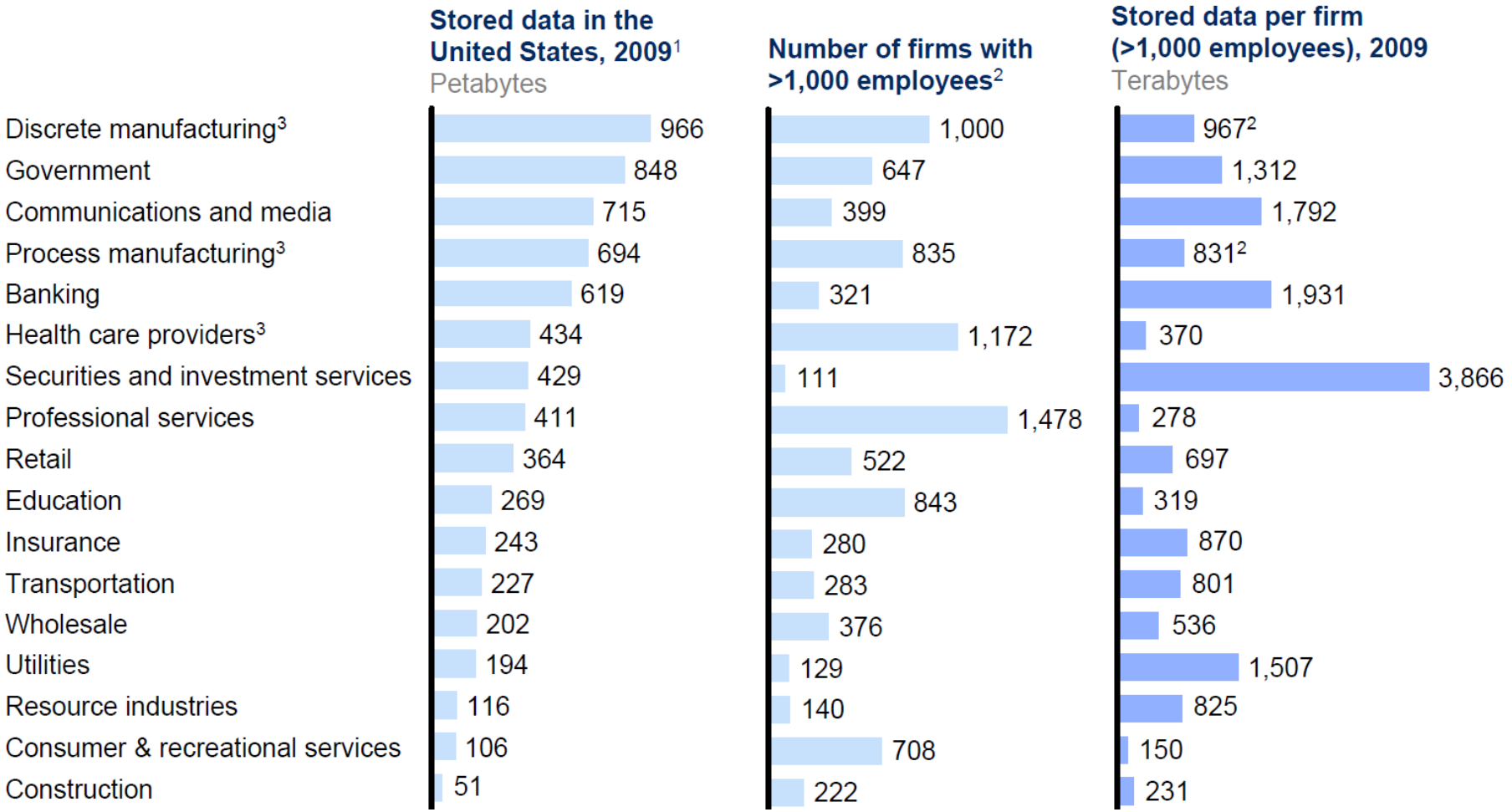
We generate more data every 2 days than we did from the dawn of civilization until 2003… and data rates are still growing—approximately 40% each year

- At Booz Allen, we've watched this trend for some time now—we call it the "**data analysis gap**"
- It's clear that **data has outstripped common analytics tools and staffing levels**
- In order to move forward, organizations must be able to analyze data on a massive scale and quickly use it to provide deeper insights, create new products, and differentiate their services



Booz Allen Hamilton and Client proprietary and business confidential

Data is exploding across all sectors of the economy

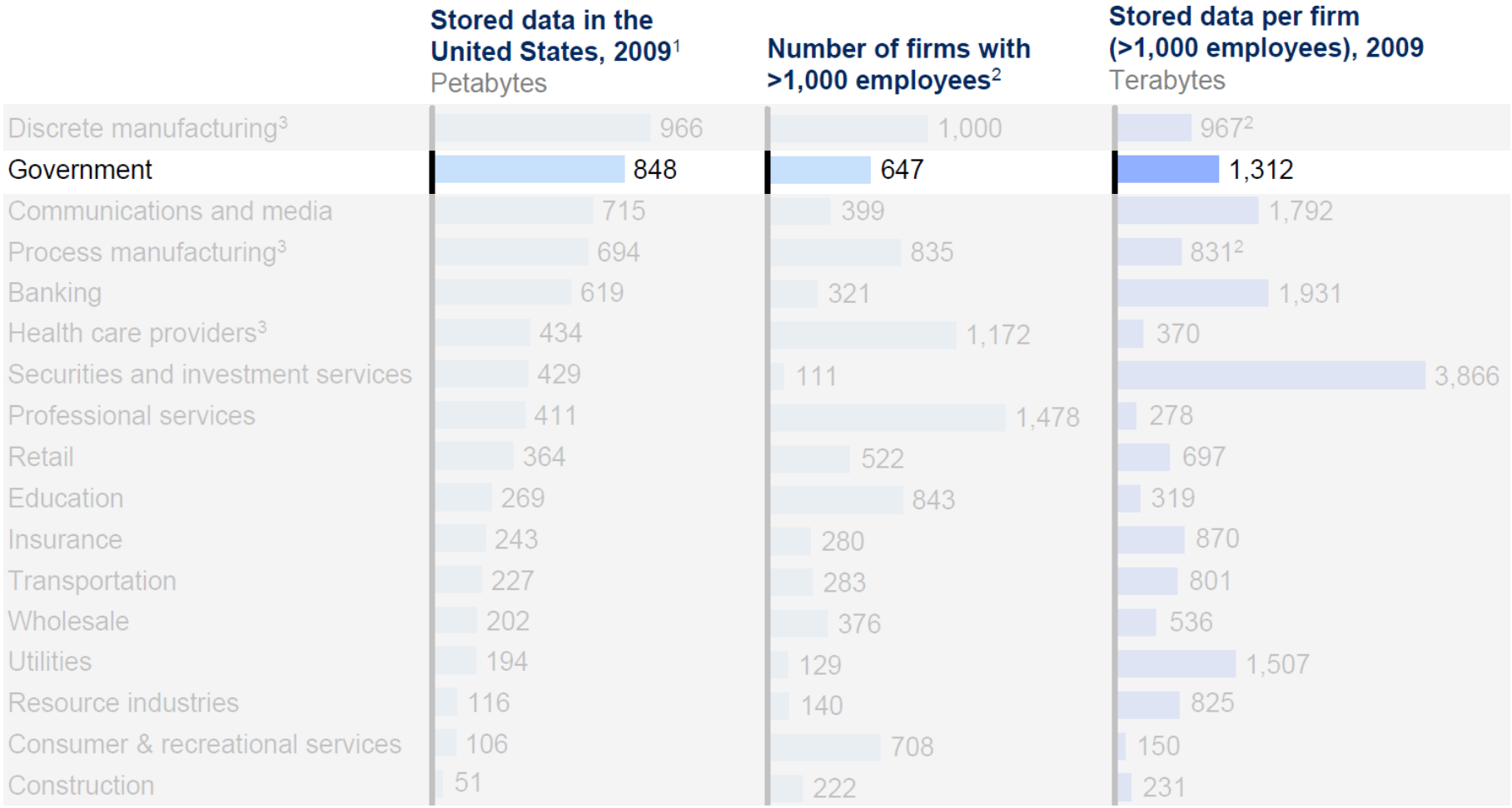


1. Storage data by sector derived from IDC.
2. Firm data split into sectors, when needed, using employment.
3. The particularly large number of firms in manufacturing and health care provider sectors make the available storage per company much smaller.

Source: IDC; US Bureau of Labor Statistics; McKinsey Global Institute analysis.

Booz Allen Hamilton and Client proprietary and business confidential

Data is exploding across all sectors of the economy

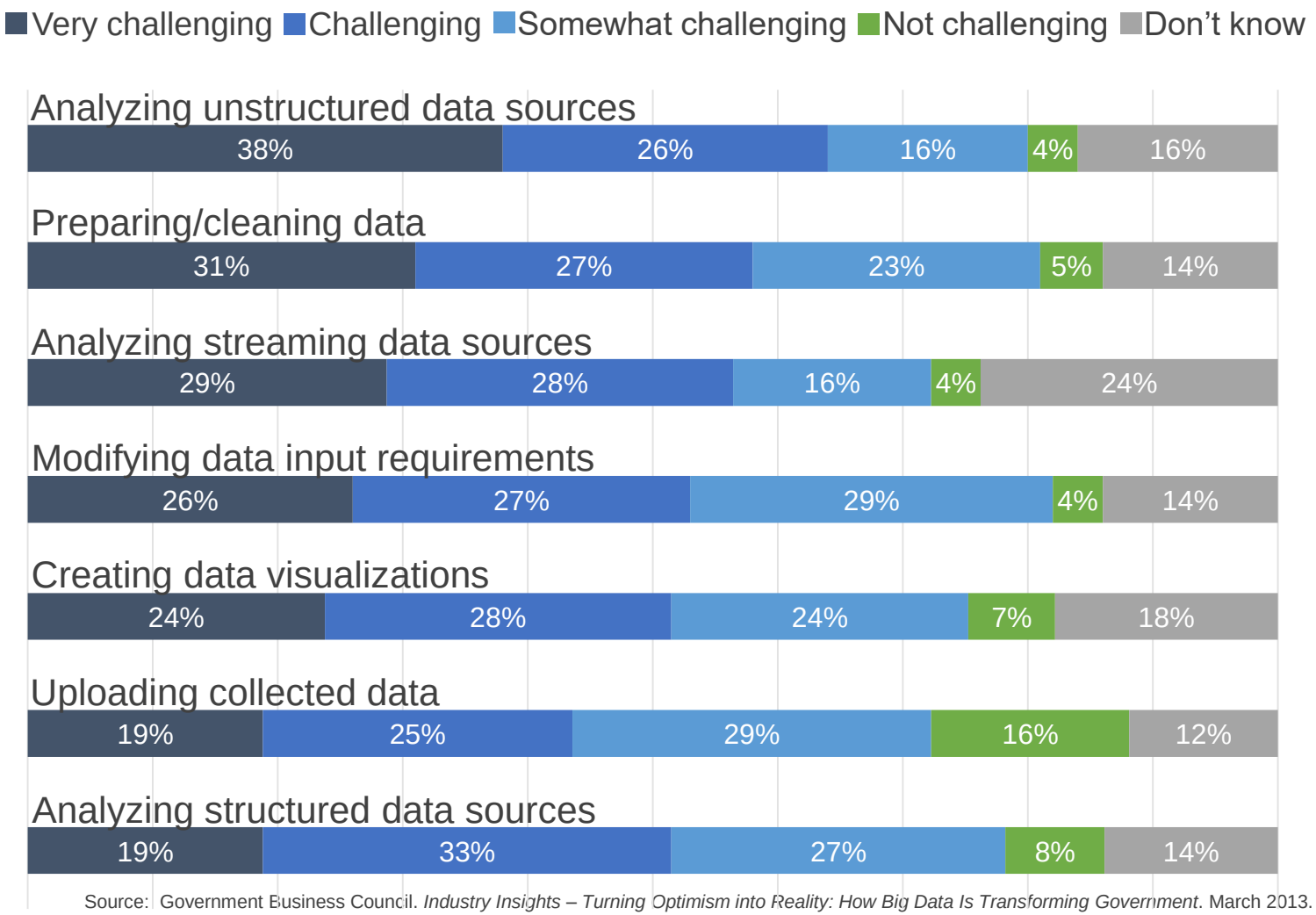


1. Storage data by sector derived from IDC.
2. Firm data split into sectors, when needed, using employment.
3. The particularly large number of firms in manufacturing and health care provider sectors make the available storage per company much smaller.

Source: IDC; US Bureau of Labor Statistics; McKinsey Global Institute analysis.

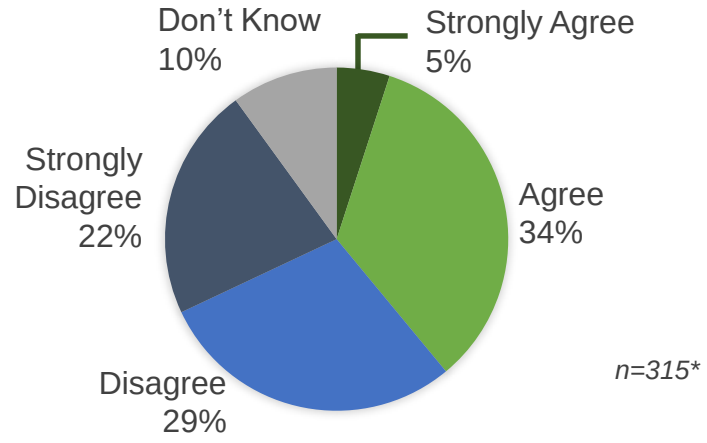
Booz Allen Hamilton and Client proprietary and business confidential

Understanding and effectively using data, especially unstructured data, is a daunting challenge for both private and public sectors

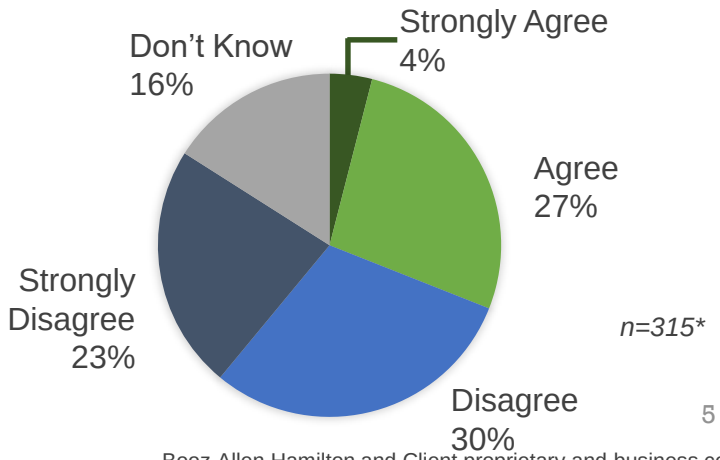


Source: Government Business Council. *Industry Insights – Turning Optimism into Reality: How Big Data Is Transforming Government*. March 2013.

My department's workforce has the technical skills to understand how to use Big Data…



My agency is fully leveraging all of the data it collects…



Booz Allen Hamilton and Client proprietary and business confidential

To harness your data, it's important to start from a services-based reference architecture that will drive your data and analytics environment

Chaotic, distributed data burdens infrastructure, impeding data analytics

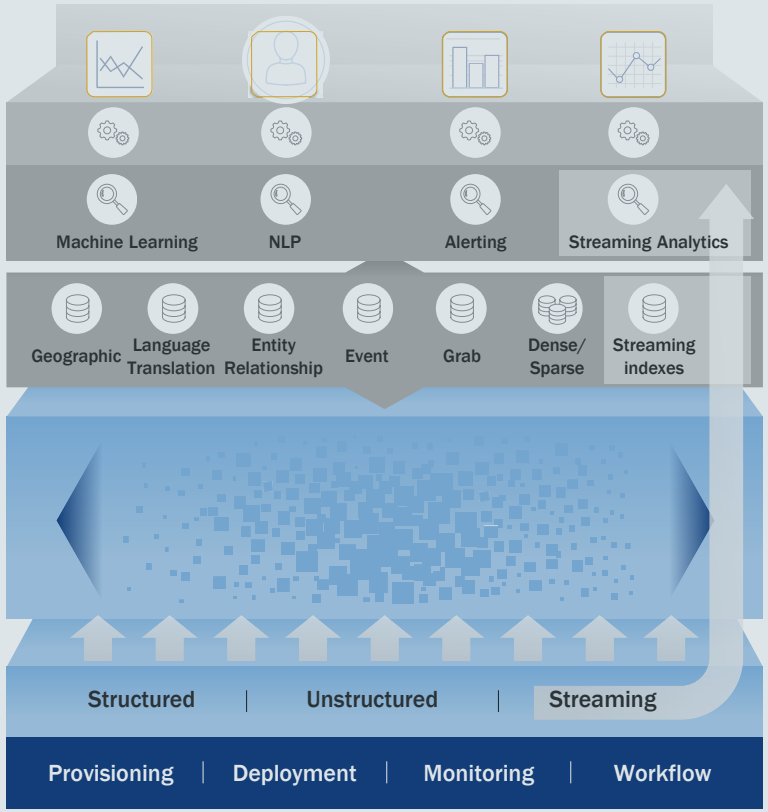
An Analytics Reference Architecture allows large-scale data processing and analytics, characterized by speed, precision, security, scalability, and cost efficiency. To achieve this, existing tools and platforms can be leveraged.

Human Insights & Actions
Enabled by customizable interfaces and visualizations of the data

Analytics and Services
Your tools for analysis, modeling, testing, and simulations

Data Abstraction
Data abstraction services eliminate the need for organizations to maintain linkages for siloed systems by creating a data abstraction layer

Infrastructure
The technology platform for storing and managing your data



Visualization, Reporting, Dashboards, and Query Interface

Services (SOA)

Analytics and Discovery

Views and Indexes

Data Lake

Metadata Tagging

Data Sources

Infrastructure/Management

Booz Allen Hamilton and Client proprietary and business confidential

We've successfully leveraged the reference architecture to build scalable data products and analytics platforms across the Federal landscape

Within GSA's Integrated Award Environment (IAE), we've **built a common services platform to host and integrate their 12 federal acquisitions systems.** A major component of the platform is the Data Lake which centralizes data across all of the systems into a single, secured environment and empowers the user to deploy analytics across all data sets. This platform **enabled non-power users to create and execute analytics/queries** across the entire environment and **integrate mission-critical acquisition and grants data into a single repository.**

Within the DoD's technology organization, **we've implemented a robust on-premise cloud-based solution that ingested, integrated, and stored 100s of TBs of data across 20 disparate systems.** The data was secured through granular-level (or cell-level) security mechanisms and provided the ability to classify data sets within a document individually for complete security controls. This new environment removed the need for individual users to interact with over 20 different silos and systems and allowed all search and dissemination to occur through a single interface.

To help the US Census Bureau with its more than 100 disparate data platforms, we've **designed and implemented a reference architecture and data lake platform** for all stakeholders. This platform **ingests, indexes, and makes searchable various data sets** including economic data, population estimates and demographic data, American Community Survey (ACS) data, Small Area Income and Poverty Estimates (SAIPE), and other types of data sets. The platform also **accommodates various unstructured data sets and was completely developed using open source technologies.**

Booz Allen Hamilton and Client proprietary and business confidential

Thank you!

Booz Allen Hamilton and Client proprietary and business confidential