

Turning Optimism into Reality: How Big Data Is Transforming Government

A Candid Survey of Federal Employees

Underwritten by:

Booz | Allen | Hamilton

delivering results that endure

About



The Research Intelligence Division

Of Government Executive Media Group



Dedicated to Advancing the Business of Government

Through analysis, insight and analytical independence



Extension of GEMG's 40 Years of Editorial Standards & Ethical Values

GBC studies influential decision-makers to produce intelligence-based analysis



Table of Contents

<u>Purpose and Methodology</u>	4
<u>Respondent Profile</u>	5
<u>Executive Summary</u>	9
<u>Research Findings</u>	11
– <u>Assessing Current Capabilities for Using Big Data</u>	11
– <u>Agency Challenges in Employing Big Data</u>	15
• <u>Interview with Dr. Joshua Sullivan</u>	15
– <u>Optimism for Big Data in Federal Agencies</u>	19
• <u>Interview with Timothy Andrews</u>	19
– <u>Making the Most of Big Data Optimism and Opportunities</u>	23
• <u>Interview with Mike Delurey</u>	23
– <u>Finding Value Through Data</u>	26
• <u>Interview with Mark Herman</u>	26
<u>Final Considerations</u>	28



Industry Insights

Purpose

The proliferation of Big Data has forced agencies to consider its great potential to revolutionize federal operations. The White House's Big Data Initiative has invested \$200 million in new research and development projects to use Big Data, and the use of Big Data in various agency activities ranging from performance tracking to budgeting has begun to take hold. As a result, many agencies are trying to train and/or hire a workforce to leverage Big Data, but in the current budget climate, new hires and investing in training courses is proving to be difficult or impossible.

In order to evaluate agency efforts to leverage Big Data and provide insight on how agencies can best capitalize on the opportunities provided by Big Data given the current budget climate, the Government Business Council, with sponsorship from Booz Allen Hamilton, undertook a comprehensive research project that surveyed federal managers.

Methodology

From January 22, 2013 to February 1, 2013, the Government Business Council deployed a anonymous survey via email to a sample of managers employed by the Department of Defense or military services. A total of 313 respondents from 27 agencies completed the survey in its entirety. For the purposes of this survey, Big Data was defined as the exponential growth of data collected, stored, and analyzed by organizations, and the challenges associated therewith, such as the storage of massive data sets or the analysis of continuous streams of data.

The following reports contains insights gleaned from analysis of survey responses, secondary research, and primary research in the form of interviews with experts in the collection, management, and analysis of Big Data.



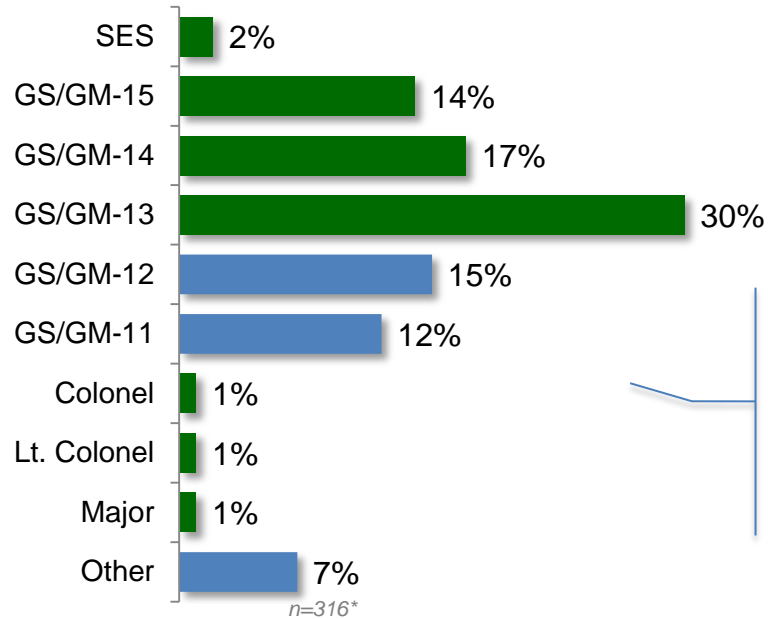
Respondent Profile

Survey Respondent Profile

- A total of 313 federal managers, from GS-11 to Senior Executive Service or equivalent grade levels completed the survey in entirety.
- Two-thirds (66 percent) of respondents are GS/GM-13 or above and 60 percent oversee at least one direct report.
- The majority of respondents oversee at least one direct report and three percent oversee at least 200 employees, making the following sample representative of the thoughts and opinions of senior federal leadership.

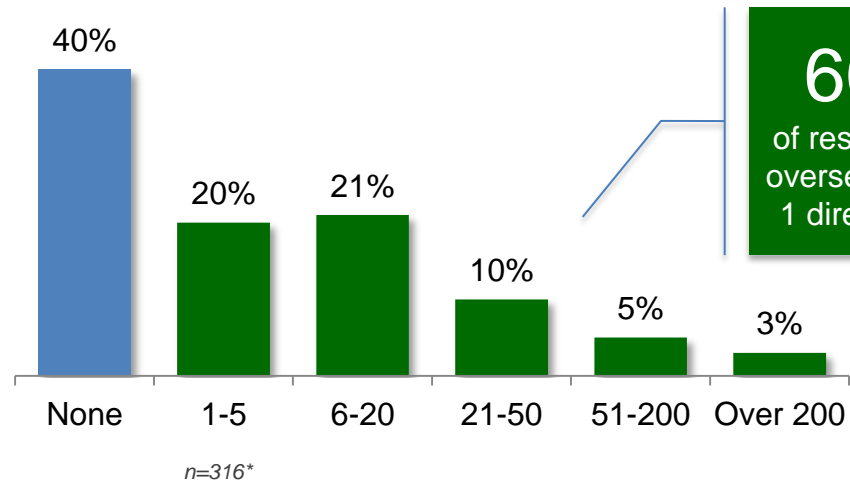
*Percentage of respondents

Job Grade/Rank



66%
of respondents
are GS/GM-13
or above

Oversees/Reports



60%
of respondents
oversee at least
1 direct report

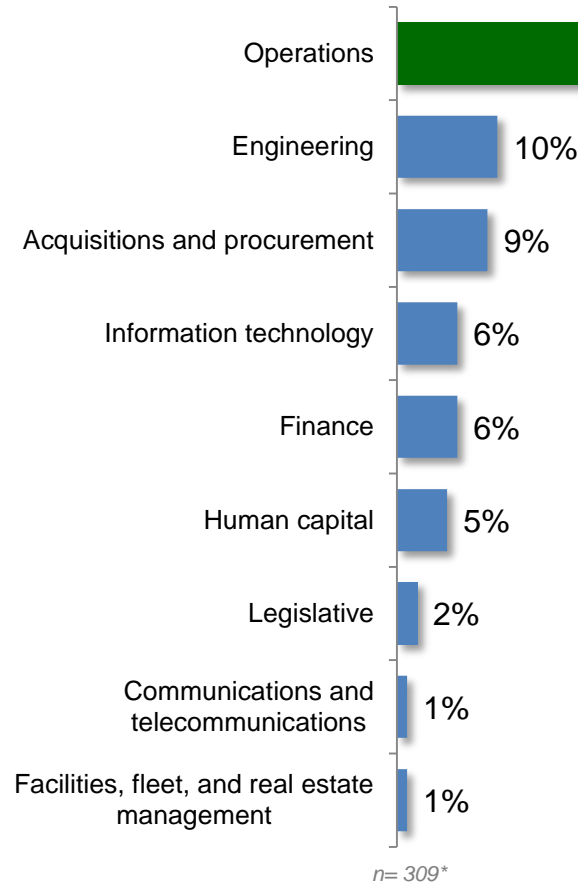
Survey Respondent Profile

- Respondents include executives from various areas of responsibility. The largest group of managers work in operations, followed by engineering and acquisitions and procurement.
- A significant portion of respondents are involved in the process of creating data analysis. IT systems operators, data analysts, and statisticians accounted for over a quarter (27 percent) of all respondents. The remainder are senior managers, specialists, and program managers (73 percent).

*Percentage of respondents

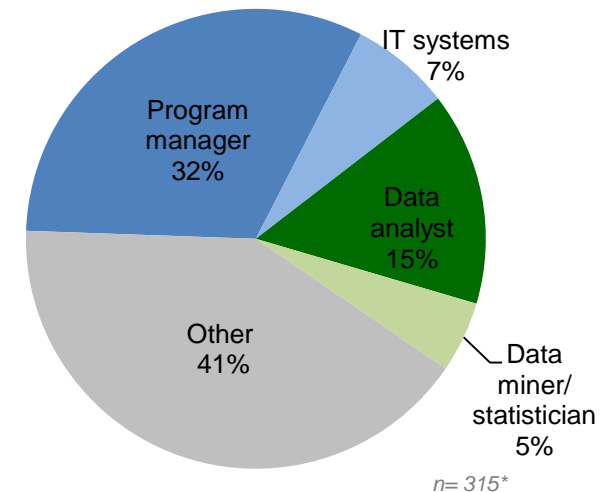


Area of Responsibility



44% of respondents indicated they were in operations.

Data Analysis Role





Respondent Profile: Agencies and Departments Represented*

- Department of the Army
- Department of Defense
- Department of the Navy
- Department of Veterans' Affairs
- Department of Homeland Security
- Department of the Treasury
- Department of the Air Force
- Department of Health and Human Services
- Other Independent Agencies
- Department of Agriculture
- General Services Administration
- Department of Transportation
- Department of Commerce
- Department of Housing and Urban Development
- Department of the Interior
- Department of Justice
- Department of Labor
- Environmental Protection Agency
- National Aeronautics and Space Administration
- Social Security Administration
- United States Marine Corps
- Department of State
- United States Agency for International Development
- Department of Energy
- Nuclear Regulatory Commission
- Small Business Administration
- United States Postal Service

**Agencies listed in order of frequency*



Executive Summary

Executive Summary

Agencies' current data capabilities could use improvement, and managers are uncertain if agencies are taking appropriate steps to course correct.

In response to the proliferation of Big Data, federal leadership has emphasized leveraging Big Data to enhance agency operations. However, managers are undecided if agencies are taking the appropriate steps to do so. Just 37 percent of managers report their agency is taking the appropriate steps, and 18 percent just don't know. Federal managers also show low levels of data proficiency, and feel the federal workforce is similarly ill-equipped to make use of Big Data. Just 18 percent of managers feel they have full professional proficiency in understanding large data sets, while 54 percent feel their agency's workforce does not have the technical skills to understand how to use Big Data. As a result, only 31 percent of respondents feel their agency is fully leveraging all of the data it collects, where 53 percent disagree.

In adopting Big Data, agencies are facing numerous challenges, and the causes for these challenges are equally diverse.

Given the limited proficiency in technical skills needed to leverage Big Data, agencies must take other routes to improve their data capabilities. Thirty-five percent of managers indicate their agency is using technology to expand their data capabilities, while 34 percent report their agency is investing their current workforce to learn new skills. Just four percent are hiring new data scientists. In adopting Big Data, agencies are facing challenges in analyzing unstructured data sources (65 percent) and preparing & cleaning data (58 percent). There are several common causes for such challenges, but the chief issues could be a lack of data visibility (i.e. managers and others do not know what data is available) (55 percent) and a lack of adequate resources (61 percent).

Despite these challenges, managers throughout government are very optimistic about Big Data's potential and expect its use to grow.

Over two-thirds of managers (69 percent) agree that Big Data has the potential to fundamentally transform federal operations beyond what is technically available right now, indicating significant optimism in government. While Big Data is currently used primarily for performance tracking and goal setting at present (53 percent), it is in the area of overall cost savings where managers expect Big Data to grow the most, increasing by 34 percent in future projections.

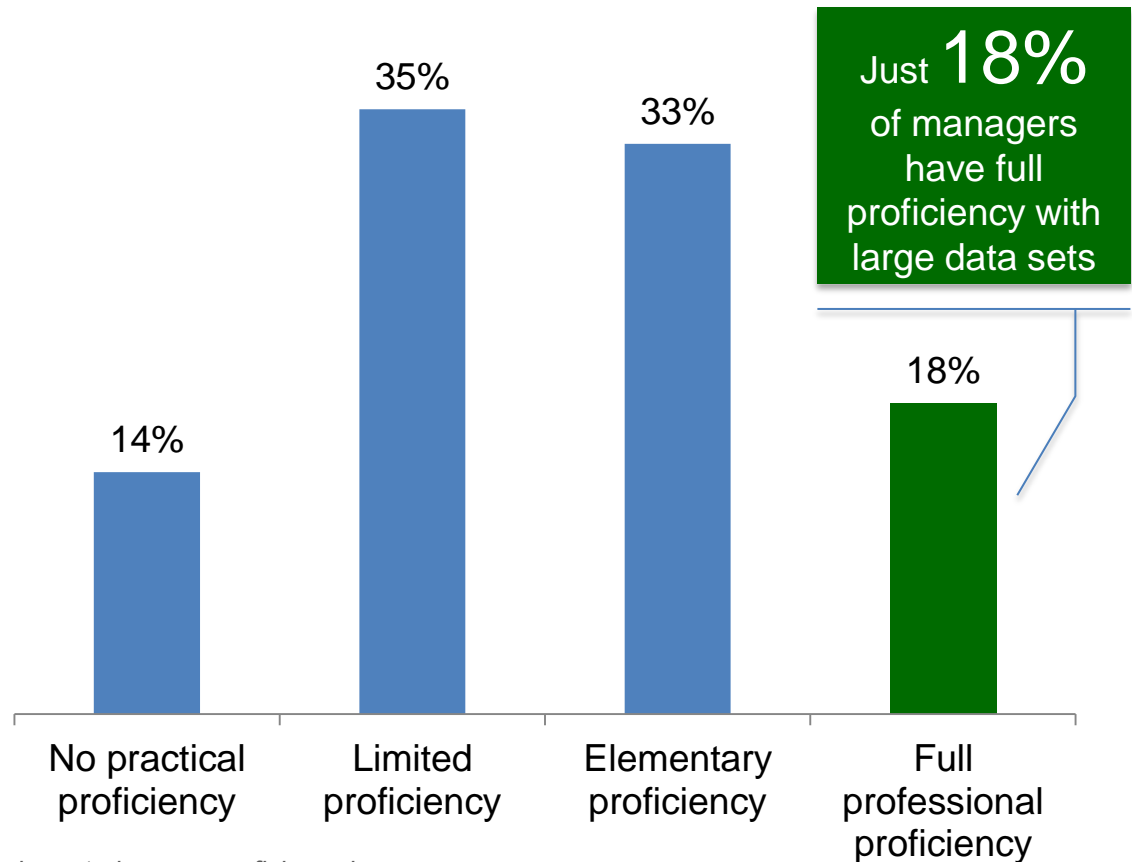


Assessing Current Capabilities for Using Big Data

Most Managers Not Proficient In Using Data; Must Rely on Analysts

- Just 18 percent of federal managers claim to have full professional proficiency in manipulating and understanding large data sets.
- While one-third claimed elementary proficiency, these managers must still rely on analysis from analysts. Therefore, managers may need to rely more on their analysts or technologies to make up for a shortfall in skills.

Data Proficiency of Federal Managers



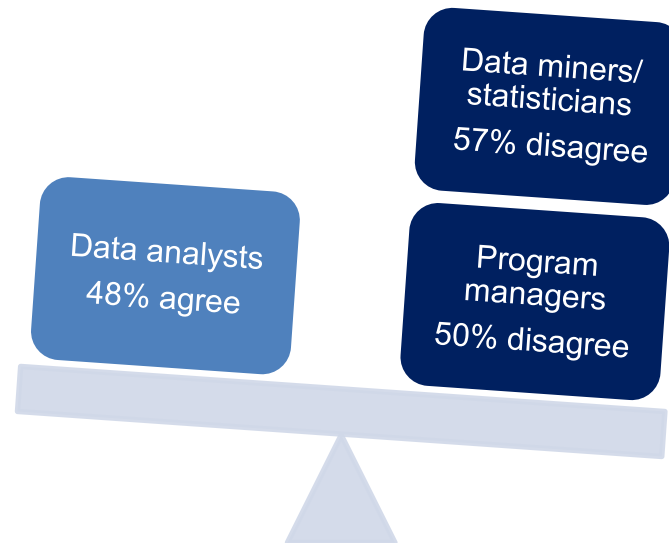
Percentage of respondents, n= 316, "How would you characterize your proficiency in understanding large data sets?"



Managers Uncertain if Agencies Taking Appropriate Steps to Leverage Big Data

- Respondents are uncertain whether or not their agency is taking the appropriate steps to leverage Big Data. Thirty-seven percent of managers overall reported that their agency was taking the appropriate steps, while 45 percent disagreed.
- Responses to this question were disaggregated by job function.
- Data analysts (48 percent) tended to have more favorable opinions of taking steps to leverage Big Data. Whereas 57 percent of data miners and statisticians tended to have unfavorable views.

Are Agencies Taking Steps To Leverage Big Data?



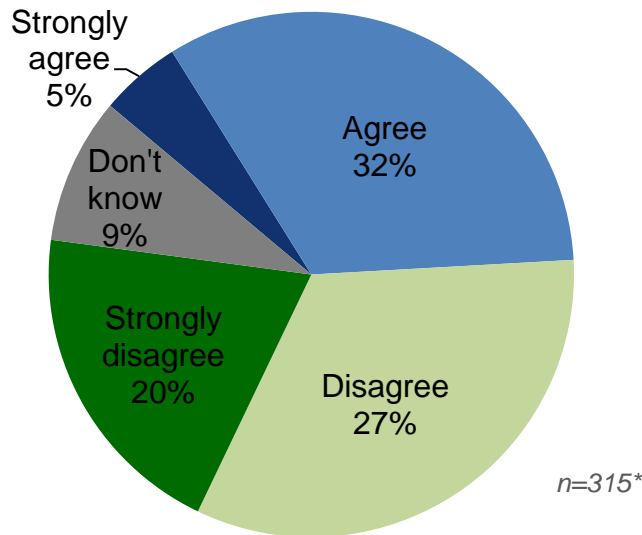
Percentage of respondents, n= 313, Likert , "I feel my agency is taking the appropriate steps to fully leverage Big Data."



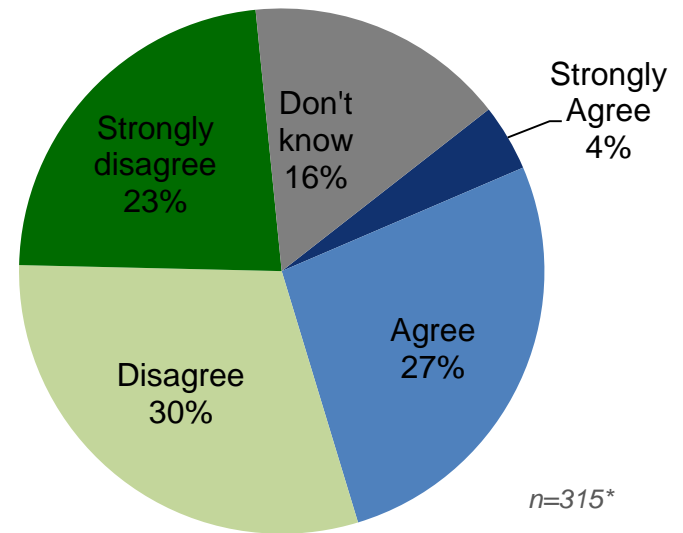
Agencies Unable to Leverage Data

- Just 37 percent of managers believe their agency's workforce has the technical skills needed to understand how to use Big Data effectively. This perception could be largely attributed to another important finding: only 31 percent of respondents believe their agency is fully leveraging all of the data it collects.

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...



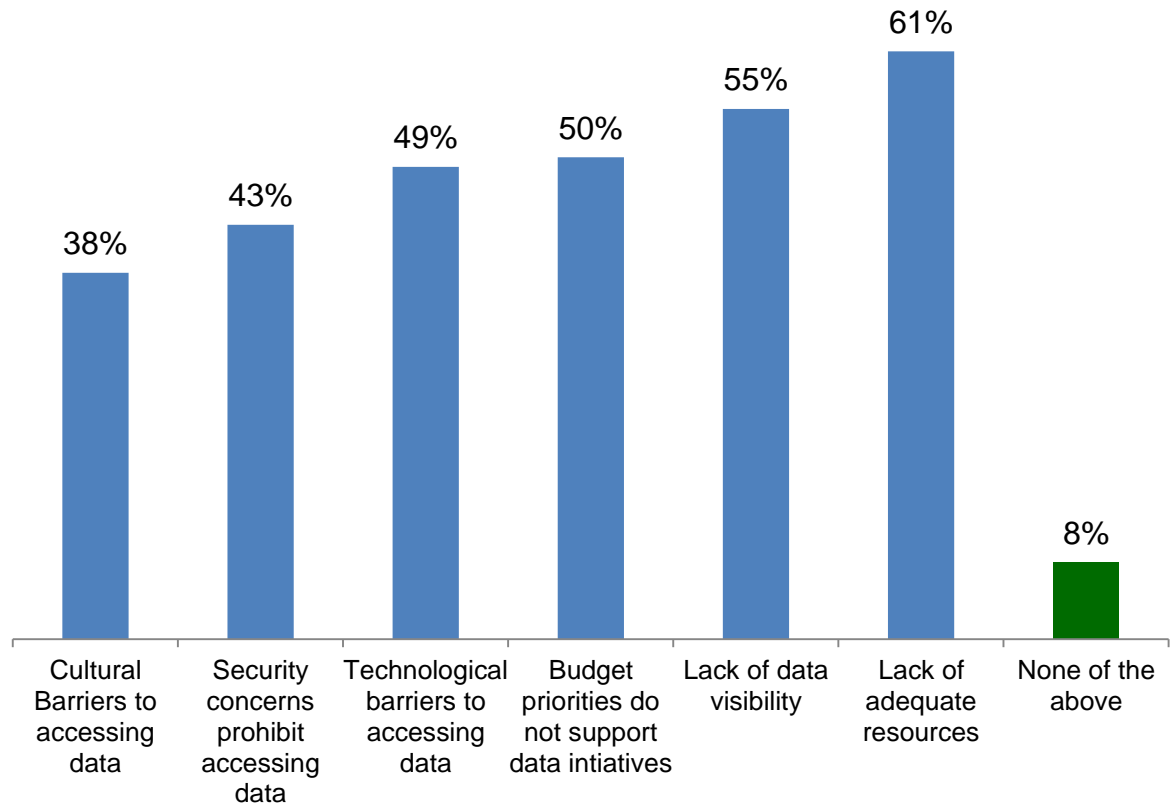
*Percentage of respondents, Likert, "My department's workforce has the technical skills to understand how to use Big Data," "I believe my agency is fully leveraging all of the data it collects."

Agency Challenges in Employing Big Data

Lack of Adequate Resources, Other Causes Hamper Agency Data Efforts

- Overall, the most common impediment in data collection & analysis is a lack of adequate resources (61 percent).
- The majority of federal managers also report facing challenges due to a lack of data visibility (55 percent), misaligned budget priorities (50 percent), and technological barriers to accessing data (49 percent).

Biggest Challenges to Overall Data Collection & Analysis



Percentage of respondents, aggregate of “very challenging” and “challenging”, n= 310, “Which of the following issues does your agency face with regard to leveraging Big Data? Please select all that apply.”

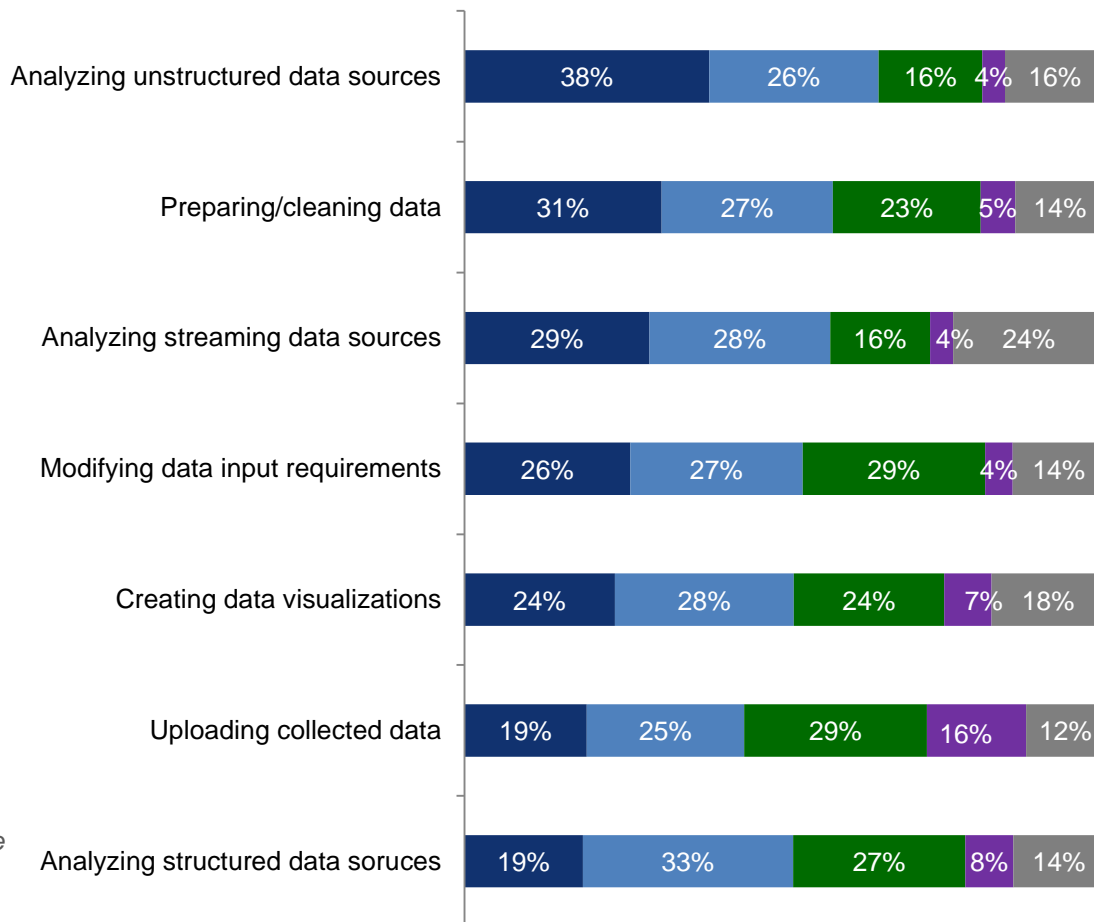


Challenges to Using Big Data

Deeper Look Into Challenges Reveals New Challenges

- After taking a more granular look at the challenges federal managers are facing with regard to Big Data, analyzing unstructured data sources nears the top of the list as 38 percent of managers report that it is “very challenging”.
- Preparing and/or cleaning data is the second most common challenge at 31 percent.

■ Very challenging ■ Challenging ■ Somewhat challenging ■ Not challenging ■ Don't know



Percentage of respondents, n varies 299, Please indicate how challenging the following steps in the data collection and analysis process are for your agency or department.





Dr. Josh Sullivan is a senior technical leader in emerging technologies, complex software development, and large-scale data analytics in a variety of domains, such as cyberspace and national intelligence. Dr. Sullivan drives the vision, strategy, and delivery of technology research and development, full life cycle system development, and data analytics.

Overcoming Big Data Challenges in Government

Big Data is having a moment of popularity in government, much like other technology fads. So what gives Big Data staying power? Why will it be relevant five years from now?

I don't think people should necessarily care so much about Big "Data" as much as Big "Analytics." It's not the data itself that will give Big Data staying power, but what we are able to learn from the data that will make it relevant five years from now. The point is to make the questions we ask of data bigger. I think federal managers will have the power to ask bigger, more ambitious questions of data, because that's where we'll find the kind of insights to transform federal operations. If we think about just Big Data, all we'll end up with are fatter stovepipes, but analytics will change that.

What are some concrete steps agencies can take to capitalize on Big Data?

There are three things that agencies really need to do to make the most of Big Data. First, they need to do a skills assessment. Most agencies are unaware if they have the right skills in their labor and contractor workforce to use data. Second, before agencies even move to Big Data, they should have a short list of the big things they want to know from their data. If agencies can't articulate a top ten list of questions that will move the needle in their agency's mission performance, then they won't even know where to begin in seeking out the right kind of technology for their mission. This list doesn't need to be exhaustive and agencies don't have to do a "big bang" technology buy, either. That leads to the third thing – this technology was designed to be implemented incrementally. Agencies can employ a variety of strategies to incrementally phase in this kind of technology, rather than buying all new parts all at once.

What do you think, given your own experience in government, is the major challenge agencies face?

A lot of government agencies don't look outside their own walls enough. And I'm not talking about looking for specific vendor solutions, but for new solutions and knowledge on the art of the possible. There's a huge ecosystem around cloud and analytics that's developing new capabilities every day. In government, too often I think they're looking to make what they have work, which is admirable, rather than looking to see what new possibilities there are. I do meet a lot of managers looking to broaden their horizons, to get outside the walls of government, but at the agency level, this kind of solution and information searching for the art of the possible is not happening. Without this kind of knowledge, agencies will continue to be limited by what they already have.

What's the future of Big Data in government?

I see two big breaking waves coming at us in the area of Big Data and analytics. One is streaming data analytics, where you can do predictive analytics on data as it's collected. When you think about the highest level cybersecurity solutions or most powerful health data analysis, those require streaming analytics. The second thing is distributed analytics. If you think about most agencies, they're building a cloud in a data center, or consolidating data center into one place. Eventually, they're going to have more than one cloud data center, and figuring out how to do analytics over data stored at multiple data centers will be the next challenge for this line of technology.

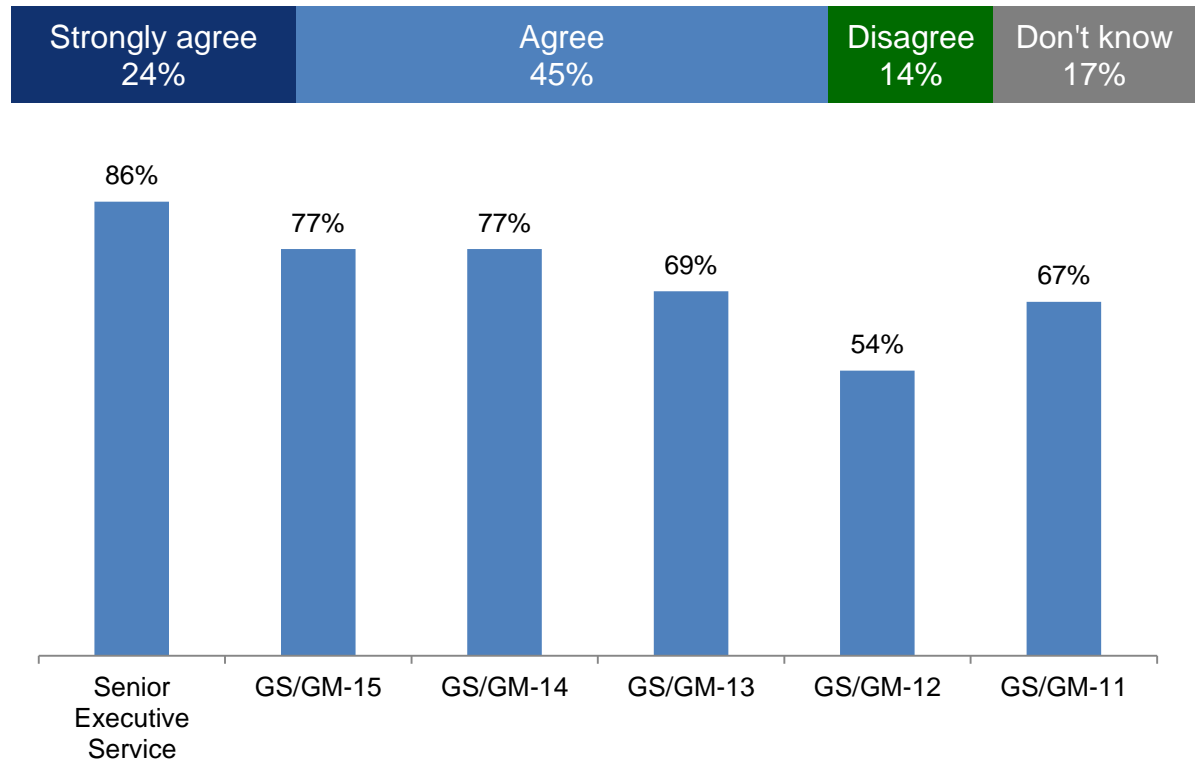


Optimism for Big Data in Federal Agencies

**Government
Optimistic About Big
Data Potential**

“Big Data has the power to fundamentally transform federal operations...”

- Managers across government show high levels of optimism for Big Data’s potential. A strong majority of respondents overall (69 percent) agree or strongly agree that Big Data can fundamentally transform federal operations beyond what is technically possible right now.
- Those managers that either agreed or strongly agreed with the stated question were disaggregated by GS ranking. Senior executive service members tended to have the most favorable opinions of the capacity of Big Data.



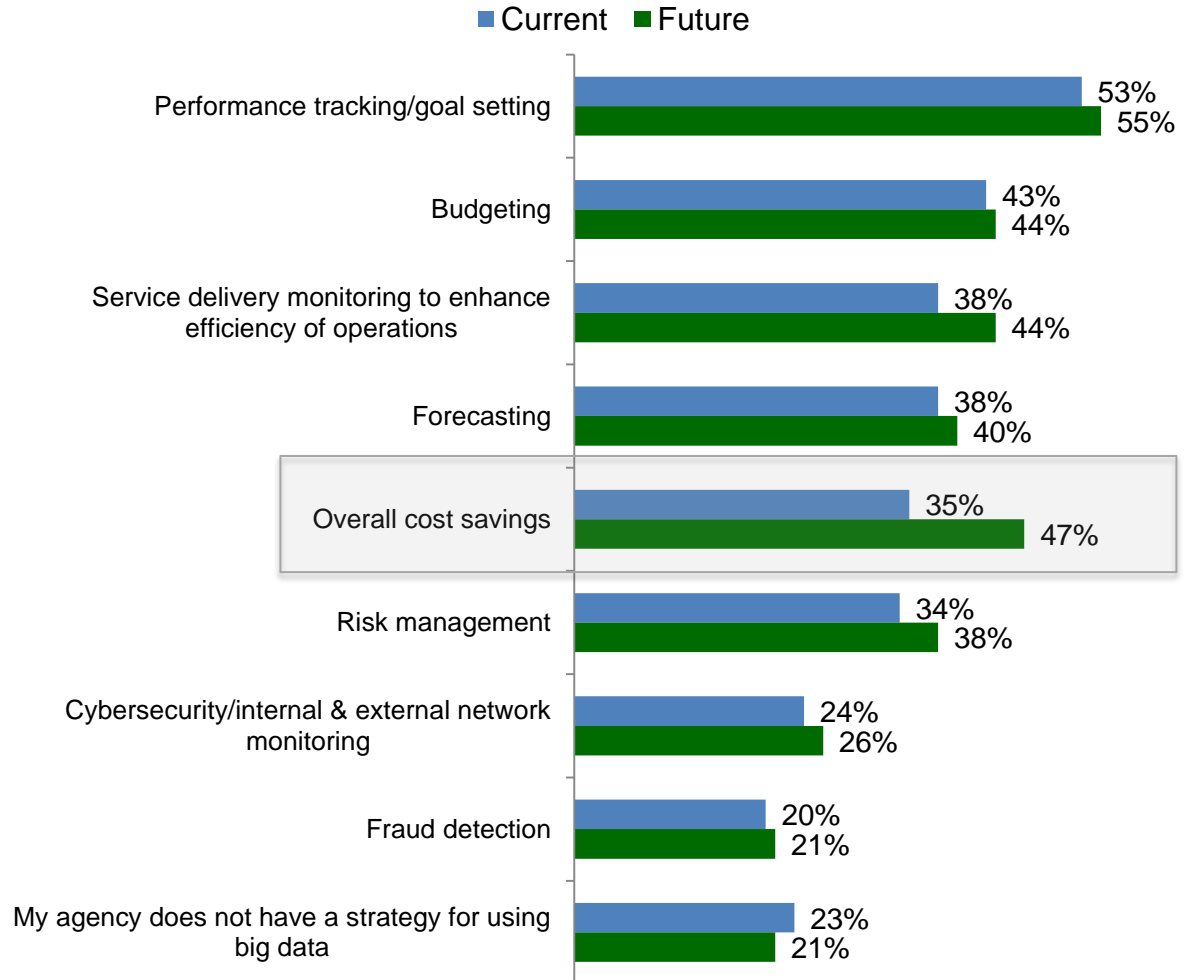
*Percentage of respondents, n=314, Likert, “The analysis of Big Data has the potential to fundamentally transform federal operations beyond what is technically available right now.”



Big Data Expected to Grow Most in Use to Create Overall Cost Savings

- We asked federal managers how their agency is currently using Big Data versus how they would like to use it in the future. Most items rated received similar scoring.
- Currently, federal managers find the value in using Big Data for performance tracking (55 percent) and see it staying relevant in the future (53 percent).
- However, Big Data is set for the largest growth in the future in the area of cost savings (%Δ=12 percent), suggesting Big Data's capacity as a cost-savings tool is growing.
- Using Big Data to enhance efficiency of operations is also set for a large growth in the future (%Δ=6 percent).

Big Data Growth Areas



Percentage of respondents, n= 311 (current), 309 (future), "For which of the following purposes does your agency or department employ Big Data?" "With regard to Big Data, which of the following goals is your agency or department trying to achieve?"





Timothy A. Andrews is a leader on the firm's health team, which provides strategy and organization, analytics, technology, and cyber security services across the payor, provider, and life sciences market segments.

Expanding Operational Capabilities through Big Data

One thing that makes Big Data exciting in the federal space is the genuine sense of optimism around it. In your opinion, what kind of potential does Big Data have in government?

Well, I'm certainly an optimist. Cloud computing is really the enabling technology behind Big Data in many ways, and with the way it is spreading, soon "the cloud" will be just like the Internet. We'll take it for granted in a few years and it'll be just another platform allowing us to do really interesting things with data. Many of the things that drove the internet, such as efficiency or cost-savings, are the same things being talked about in Big Data, which is of course built upon the cloud. And though maybe not at the same scale as the Internet, Big Data can really transform what agencies are capable of doing, and that will foster widespread adoption.

You say Big Data will transform operations – how so?

What's so transformational about Big Data is that it fundamentally alters the way we can solve problems. Before, we always had to decide what answers we wanted to know, to a very granular level, from the data. With Big Data, we can start with questions first from a much more exploratory stance. We can ask a lot of really challenging questions and get answers from all of the data, even new data that is uploaded later on. In many operations, you don't know what data and how much data you're going to have by the end. But now you can ask new questions as you go along, and ask that of all the data. One good example is trying to collect data on flu shot programs. When someone gets a flu shot at a CVS or other private clinic, then goes to their doctor, the doctor will often make a note that the patient got a flu shot in an open field. But the EHR system will often not report the flu shot since most systems don't parse text from unstructured fields. So trying to analyze the number of flu shots is problematic without using advanced analytics like free text. Thus a better

approach might be to just put the EHR data into a data lake and then use a variety of techniques to try and discover a full medical history by analyzing the structured and unstructured text.

What steps can agencies take towards finding ways to put their data to use in operations?

The big first step is to create a data strategy, and within that, creating a data stewardship program. Agencies are realizing the value of collaboration and data sharing, but technical and organizational issues such as data visibility will limit those efforts. Organization such as federal agencies are massive sprawling entities, and because of the way funding works, a lot of programs build independent systems, so finding data can be near impossible, even if you know it exists. An empowered data steward can facilitate collaboration and improve data visibility. Also, in many cases, such as in healthcare, data is highly sensitive, and then there are also issues of secondary use of data, such as healthcare data used to identify wrongdoing. These kinds of governance questions are the first obstacle agencies face in employing data in operations, and the best way to start solving them is with a data strategy.

Have you seen any areas where Big Data has been employed to improve operations?

Though it's still very nascent, Big Data has seen some impressive growth in the area of Health Information Exchange (HIE) for example. There are lots of data challenges in HIE programs, such as integrating clinical data with billing data, or keeping data secure and confidential while still trying to analyze it to improve service delivery. With Big Data technologies, HIE programs will be able to analyze and share data without compromising patient data and privacy, enabling a whole new level of capabilities and improvements in service delivery.



Making the Most of Big Data Optimism and Opportunities

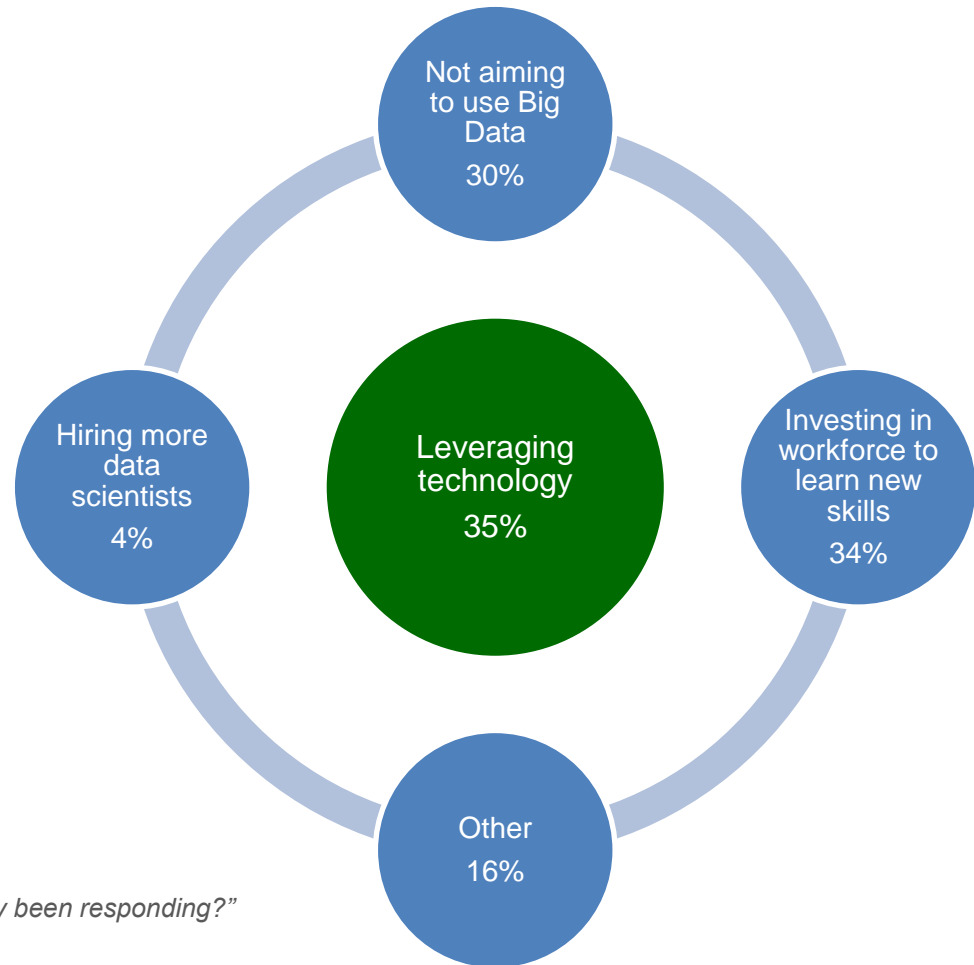
Industry Insights

**Leveraging Technology
Most Common Agency
Response to Rise of
Big Data**

- Just over a third (35 percent) of managers indicated that their agency is leveraging technology to use Big Data more effectively, while 34 percent are investing in their current workforce to learn new skills.
- Just four percent are hiring more data scientists, while just under a third (30 percent) report their agency is not aiming to use Big Data. Under tightening budgets, federal managers may find solutions in technology to account for the growing potential of Big Data.

Percentage of respondents, n= 315, "How has your agency been responding?"

**Agency Response to Proliferation of Big
Data**



Mike Delurey concentrates primarily in the areas of policy analysis support, critical infrastructure protection, strategic risk management, complex systems analysis, homeland security, mission assurance, information assurance and cyber security policy. His focus is on the intersection of technology and policy issues.



Capitalizing on Big Data's Potential with a Data Lake

Big Data implies a lot of different technologies, and one Big Data technology that is gaining steam is using a “data lake.” In laymen’s terms, what is a “data lake”?

One of the easiest ways to understand a data lake is to look at what it’s being compared to. The data that is being accessed right now is usable because it’s stored much like a library, where a book is placed in a certain location, and it can be found with a catalogue number that denotes a specific location. But when you have massive amounts of data, storing data like a library is difficult and inefficient, so you need a system that can locate data without knowing its specific location. That’s what’s drove the move to a data lake. With it, we can access and store data without having to put in a specific location, and that enables us to not only store massive amounts of data, but also to run complex analytics over all the data.

And how would those benefits translate to government? What kind of potential does the data lake have?

One of the major challenges creating inefficiencies in government are the limitations put on agencies to collaborate. Short history lesson: in 1934, Congress passed the Economy Act of 1934, which creates operating islands by separating budgets. Agencies can’t easily cross fund initiatives which may benefit from collaboration without Congressional approval because of this act. What a data lake can do however, is lower the silo walls between agencies and connect them at the data level. Because it’s low-cost and scalable, the data lake is viable option for agencies to at least share their data efficiently, even if they can’t operate programs jointly, and to make the data actionable.

Can you give any specific examples of how a data lake might improve agency operations?

My own specialty is in the area of security and intelligence, and what the

defense and intelligence communities have spent the past twenty years trying to improve is the “observe, orient, decide, act” loop (OODA). It’s essentially the decision-making process. The faster you can observe, orient, decide, and act the better, and the stronger your advantage over adversaries will be. Arguably the most important part of the OODA loop is orienting yourself, because it affects how you observe, the decisions you make, and how you act. Under current technological constraints, orientation is limited by preset hypotheses. But a data lake lets agencies approach data without preset hypotheses they’re trying to prove and to ask all new questions. This means that not only are you getting better orientation from the data, but you’re doing it in much less time, speeding up your operational tempo.

Given this potential to improve operations, what do agencies need to do make something like a data lake work for them?

The most powerful tool that agencies have to move toward something like this is to create a strategy, a data strategy. Defining a strategy is one of the things institutions are often challenged to do so clarifying a strategy is the first step. Without a data strategy, you can’t understand what is the art of the possible and it’s difficult to orient the benefits to the mission requirements. With a strategy however, it’s possible to work through the impediments and begin to realize how a move will actually be accomplished.



Finding Value Through Data



Mark Herman is a senior leader of the firm's Cloud-based Services business. He has more than 30 years of experience designing and developing wargames on a variety of strategic and boardroom topics for the U.S. Department of Defense and Fortune 500 companies..

The Executive Case for a Data Lake

Big Data is creating a transformation in government. Can you give us any examples of what this transformation might look like?

Two of the big areas I've followed are the regulatory environment and the waste, fraud, and abuse dimension. Since 2007-2008 financial scandals, Congress rolled many new regulations that banks must comply with. I've worked with a financial client who estimated that new regulations would cost an additional \$600 million dollars a year in IT. With a data lake and cloud platform, the client would gain a whole new level of capabilities to apply new regulations to product work streams, prove compliance with those regulations in real time, and as a result, that client will probably spend half as much as they would have to comply with regulations.

From a 10,000 foot view, what efficiencies can be gleaned from a data lake?

What makes a data lake so valuable to someone like a CIO is actually the human element. Agencies are creating and collecting so much data everyday, it's impossible to make use of it all because it has to be structured to be analyzed. And that usually entails a human being involved in structuring that data. A data lake automates the ingestion process and removes the need for structure. This frees up a significant piece of the workforce that can save costs and allow a CIO or CEO to utilize that talent for critical thinking functions or high level data analysis, areas where you want human, not machine, eyes. Big Data doesn't have a silver bullet answer to all problems, but including a machine element that frees up labor to spend more time doing more important things will improve operations and cut costs.

When can an agency make the most of a transition to a data lake and cloud analytics?

The amount of money an agency saves really depends on its IT liquidity—it can be the difference between 10 percent and 40 percent savings. If an agency has just spent a large sum of money on new systems, it probably won't save a lot of money moving to the cloud. If an agency has a lot infrastructure set in place already, it can still save money by setting a strategy to phase in new technology as the old infrastructure reaches the end of its life cycle. And those organizations struggling with hardware that is on its last legs will obviously save the most. But regardless of where an agency's IT infrastructure is, it can save money. Different situations just require different approaches and different timelines to realize cost savings.

From the lens of advancing the mission, what can a data lake offer?

Government doesn't really focus as much on profit as it does on mission improvement. With a data lake, the productivity and the missions of a labor force change, from an assembly line process to a higher cognitive process. A data lake frees up analysts to spend more time doing analysis rather than mundane data management, so the information senior leaders are provided is much improved. Senior leaders have to make decisions for an organization, regardless of whether there is good, bad, or no data to use for the decision making. A data lake ultimately improves the quality and timing of critical operational data. So really, a data lake affects processes and decisions across an enterprise, and that improves mission execution.



Final Considerations

Agency leadership is optimistic about future cost savings resulting from Big Data technologies.

Federal managers were asked what their current agency is doing to leverage Big Data versus what their agency goal is. Thirty-five percent of respondents indicated that their agency is currently looking to use Big Data to create overall cost savings and 47 percent said it was a future goal. Agencies see the value in Big Data and see its value in the future. Current capabilities may limit agency goals, agencies should look to other agencies and organizations for examples on how Big Data has been used successful for cost savings and/or cost avoidances.

Challenges to Big Data implementation vary, but some trends appear.

Thirty-eight percent of federal leaders indicated that analyzing unstructured data sources was “very challenging.” When data was aggregated between “very challenging” and “challenging” items, the biggest challenge was the lack of adequate resources. Despite these different findings, federal leaders may lack or not know of the resources that could assist them in leveraging data. New technologies that are customized to meet agency needs may hold the answer.

To capitalize on Big Data opportunities, agencies will need a Big Data strategy.

After analysis of the results, several experts conclude that a Big Data strategy is necessary for finding value in Big Data. Before adopting a technology solution or bolstering the analytic workforce, agencies need to identify the key questions and challenges they have that Big Data could address. Only after going through that process and creating a holistic strategy, can the value of Big Data be realized.



Contact

Contact

Dana Grinshpan

Research Manager

Government Business Council

Tel 202.266.7862

[**dgrinshpan@govexec.com**](mailto:dgrinshpan@govexec.com)

<http://www.govexec.com/GBC>

OnTwitter:

@GovBizCouncil

Underwritten By:

Booz | Allen | Hamilton

delivering results that endure

About Booz Allen Hamilton

Booz Allen Hamilton has been at the forefront of strategy and technology consulting for nearly a century. Today, the firm provides services primarily to the US government in defense, intelligence, and civil markets, and to major corporations, institutions, and not-for-profit organizations. Booz Allen offers clients deep functional knowledge spanning strategy and organization, engineering and operations, technology, and analytics—which it combines with specialized expertise in clients’ mission and domain areas to help solve their toughest problems.

Booz Allen is headquartered in McLean, Virginia, employs approximately 25,000 people, and had revenue of \$5.86 billion for the 12 months ended March 31, 2012. To learn more, visit www.boozallen.com. (NYSE: BAH)

Turning Optimism Into Reality: How Big Data Is Transforming Government

A Candid Survey of Federal Employees

Underwritten by:

Booz | Allen | Hamilton

delivering results that endure