Addressing government challenges with big data analytics
Contents

2 Introduction
4 How big data analytics deliver greater value to government
6 Successfully deploying big data analytics
7 The IBM big data platform

Introduction
Governments worldwide have a wide range of responsibilities across national, regional and local jurisdictions. However, today's environment presents significant economic and budgetary pressures for public-sector organizations, as well as increasing demands to deliver better levels of service to citizens and businesses. Not only are governments expected to become more responsive to the public's concerns and provide transparent accountability and open access to decision-making processes, but they are also required to address growing threats to public safety, national security and the environment.

No matter where they are located or what range of constituents they serve, all government organizations are looking for ways to become more efficient, reduce or maintain costs, offer better services and increase responsiveness. Many are exploring ways to leverage big data analytics to better address challenges and improve operational efficiencies and services.

Big data, which is defined as large volumes and wide varieties of data that are sometimes delivered at high velocities, is already being collected by many government organizations—and new data is being generated at rapidly accelerating rates. The ability to perform analytics on big data enables government organizations to improve existing processes and operations, and engage in entirely new types of analyses that weren't possible before.

The types of big data that governments may want to analyze include records about citizens, businesses and transactions stored in core systems and data warehouses; text documents including email, blogs, social media, chat-room exchanges or medical notes; and other data such as videos, broadcasts, images, sensors, and geospatial and weather information.
The data varies in format from structured to unstructured, and is either at rest or in motion. Traditional structured data, which includes information such as tax, crime or motor vehicle records, is usually stored in legacy systems or data warehouses. Unstructured data includes social media, email, videos, images and so on. Data at rest already resides in a database or data warehouse somewhere, while sensors, satellites, the Internet, broadcasts and other streaming data sources continuously generate data in motion.

As shown in Figure 1, big data analytics has the potential to greatly benefit most areas of government by enabling:

- Richer holistic views of citizens and other entities, as well as their associated activities and events
- Zero-latency operations with immediate access to data and insights
- Rapid delivery of information including insights about fraud or risk
- Immediate retrieval of streaming data from instrumented assets

This white paper provides an overview of how big data can benefit government agencies, including examples of how some IBM government clients are benefiting from big data analytics, what agencies need to consider when adopting new big data analytics solutions and how IBM can help.

Figure 1: By harnessing big data, government agencies can perform a host of tasks, including responding to public safety or environmental events quickly, supporting decisions with both historical and real-time data analysis, and facilitating timely investigation of crimes, fraud and other incidents.
How big data analytics deliver greater value to government

Big data analytics can deliver many benefits to government organizations and programs—from increasing efficiencies through the delivery of online services and information to citizens, to ensuring a nation or region is protected from terrorist attacks. Key areas where government organizations have achieved initial success with big data analytics include the detection of fraud for social programs and tax collection, and the prevention and prediction of threats and crimes.

Fraud detection

Fraud, abuse and errors can cost government agencies billions of dollars each year. Two areas where governments have much to gain from reducing fraud exposure are social programs and taxes. More accurate and rapid detection of potential problems can help governments reduce the costs of social programs and increase tax collections and revenues. By collecting and analyzing data from a wide variety of structured and unstructured sources, such as medical and caseworker records and notes, financial transactions, watchlists, death records, crime data, social media, geospatial information, videos, email and other data sources, big data analytics can help to automatically identify hidden relationships and activities that may point to fraud or errors. Once potential problems are detected, agency personnel can investigate specific cases and make a final determination as to whether fraud, abuse or errors are actually occurring.

Social programs: Social programs provide citizens with much-needed services in areas such as employment, public health, workers compensation, child welfare, and family, child and income support. Spending on social benefits is one of the largest expenses of most governments, and it is not uncommon for large amounts of money to be paid out in benefits that are lost due to fraud, duplicate payments, overpayments and errors.

Big data analytics can help prevent unnecessary payments by providing agencies with the additional insights they need to detect fraud and errors before payments are made, reduce the time required for caseworkers and analysts to manually investigate potential fraud and errors, and give caseworkers a clearer view of their cases.

• A large national medical and social benefits organization used big data analytics, including ad hoc queries of more than 70 data sources, to reduce the time it took to conduct analyses from weeks to just four hours. More than USD140 million in improper payments were immediately identified, including payments made to thousands of dead people.
• A regional social programs agency deployed big data analytics to help eliminate waste, fraud and redundancy by identifying relationships between benefit recipients and programs. Once the system was in place, the agency was able to immediately reduce waste by USD11 million.

Tax agencies: Tax agencies need to minimize tax gaps and increase revenue collection by ensuring that all entities pay their required share, refunds are issued only to those who legitimately qualify for them, and audits are performed on those most likely to be committing fraud, underreporting income or participating in other tax evasion schemes. Big data
analytics can help tax agencies more accurately determine who should be investigated for fraud or denied refunds by detecting new deception tactics, uncovering multiple identities and identifying suspicious behavior.

- A national tax authority wanted to reduce the time it took to complete audits. The organization deployed big data analytics to enable employees to complete complex audits more efficiently. The system generates reports approximately 12 times faster than before to provide near-real-time access to information and provides the ability to rapidly perform ad hoc queries.

- A regional tax authority was seeking ways to close its tax gap. The agency, which processes more than 24 million business and personal tax returns and collects more than USD90 billion in tax revenues annually, deployed big data analytics to help it better determine which cases should be audited and investigated before refunds were issued. Initial results in the first year included a seven percent increase in collected tax revenue and the recovery of USD83 million in delinquent taxes.

Predicting and preventing threats and crimes
Two primary governmental functions are national security and public safety. Predicting potential threats and crimes before they happen and preventing them from occurring can significantly lower risks and improve public safety and national security. Big data analytics can help national security and law enforcement agencies improve intelligence by identifying threats and crimes before they occur, finding critical information faster, detecting associations between people and activities, improving the accuracy of threat and crime analysis, enabling information sharing and collaboration between investigative organizations, protecting sensitive facilities from attack and thwarting growing cybersecurity risks.

For both national security and law enforcement agencies, time is of the essence. They need to be able to combine vast stores of structured and unstructured historical data, with real-time streaming data from sensors, satellites, aerial cameras, video cameras, social media, voice, chat rooms and other sources, and immediately aggregate all sources to create a complete, holistic view of an entity or situation. Analytics must be automatically performed on aggregated data to provide analysts and investigators with highly accurate, real-time identity, location and relationship insights and intelligence about targets, areas of interest and patterns of life.

- A high-security government facility is using big data analytics to protect its perimeters and borders from potential intrusions and threats. The system continuously collects, processes and analyzes up to 1.6 GB of real-time, streaming, digital acoustic sensor data and images from video and airborne surveillance systems and provides personnel with a holistic view of potential threats, issues or nonissues. With better intelligence, the facility can more rapidly identify and classify potential security threats, above and below ground, and take appropriate action even when threats are miles away.

- The police department in one of the largest cities in the world is using big data analytics to redefine how information is used to fight crime. The system analyzes and integrates all information assets in real time across structured and unstructured data sources to facilitate faster and more appropriate responses. New comprehensive insights can now reach detectives while they are on the scene, rapidly identify repeat offenders and quickly detect trends that enable resources to be deployed faster and more efficiently.
Successfully deploying big data analytics

As shown in the previous use cases, big data inverts the traditional analytics model. Historically, business users determined in advance what questions they wanted to ask based on data they knew their organizations were already collecting or could access. Traditional analytics systems were built and structured around the data required to answer those specific questions. In contrast, big data analytics systems are designed to enable creative discovery based on wide varieties and large volumes of data, some of which users may not anticipate in advance. These systems allow for iterative and exploratory analytics that provide users with much more flexibility to explore options and ask a wider range of questions.

So how can public-sector agencies take advantage of big data analytics? The TechAmerica Foundation Big Data Commission addressed that question in its October 2012 report entitled “Demystifying Big Data: A Practical Guide to Transforming the Business of Government.” The report was created to provide “the [US] government’s senior policy and decision makers with a comprehensive roadmap to using Big Data to better serve the American people,”¹ but the recommendations put forth in the report are just as relevant to other government organizations worldwide. IBM was a major contributor to the report, and Steven A. Mills, senior vice president and group executive, software and systems at IBM, co-chairs the Big Data Commission.

In its report, the Big Data Commission recommends agencies interested in exploring big data opportunities follow a roadmap with five steps: define, assess, plan, execute and review.² These steps, detailed below, are the same ones that IBM suggests government agencies take when deploying big data analytics.

1. Define the use case or set of use cases that will be part of the initial big data deployment, key business objectives, and specific mission requirements and strategies. In most cases, the goals should be to start with projects that can deliver the greatest value in the shortest timeframe and plan for how to expand over time.

2. Assess the agency’s currently available data and technical capabilities against the data and determine what additional data or capabilities are required to satisfy the defined set of business and mission requirements.

3. Plan for deployment by determining what big data entry points the project will emphasize (volume, velocity or variety of data), what deployment pattern is best to support the entry points, which existing IT investments can be integrated rather than building new systems, and how security, privacy and policy requirements will be fulfilled. The technical architecture should be designed to support the current project, but it also should be capable of scaling to support broader goals and use cases. A comprehensive information foundation should be put in place that is capable of supplying accurate and secure information, insight and action.

4. Execute the planned project by deploying elements of a big data platform based on the recommended plan and architecture. Keep future projects in mind to help ensure the platform is flexible and able to expand to meet eventual volume, velocity and variety demands along with growth requirements.

5. Review and report on roadmap steps throughout the process, and adjust plans based on findings and lessons learned by an iterative deployment process. Review plans to ensure they meet security, privacy and governance policies. Evaluate architecture and integration technologies for problems or opportunities to leverage capabilities across departments and organizations. Assess cost and timeline results and measure return on investment to assist with planning future projects and reporting on fiscal responsibility.

To help ensure the success of big analytics deployments, the IBM® big data platform offers a full range of big data technologies and services, from consulting and technology to experienced professionals who can assist with implementation and testing.
The IBM big data platform
IBM offers a comprehensive collection of best-of-breed technologies and assessment and implementation services that help organizations integrate data from disparate sources, analyze big data in real time, help anticipate future outcomes and rapidly generate insights to capitalize on new opportunities (see Figure 2).

The platform includes a broad range of components that enable government organizations to harness the power of big data analytics, including:

- An enterprise-ready version of the Apache Hadoop engine to cost-effectively process and analyze massive volumes of structured or unstructured data

Figure 2: The IBM big data platform helps agencies address a variety of data-driven processes, as well as improve efficiency and deliver deeper, more timely insights.
• A streaming data platform that enables the application of analytics directly on data in motion in sub-millisecond response times to take action in real time
• High-performance data warehouse appliances for complex analytic processing
• Text and content analytics to uncover hidden meaning and insight in unstructured information such as case notes, social media and more
• A navigation tool to find, understand and navigate big data while leaving that data in place
• High-volume information integration applications to ensure data is clean, trustworthy, protected and secure

The leading IBM products that deliver these capabilities include:

• **IBM InfoSphere® BigInsights** provides an integrated solution for analyzing hundreds of terabytes, petabytes or more of raw data derived from an ever-growing variety of sources.
• **IBM InfoSphere Streams** provides a state-of-the-art computing platform that can help companies turn burgeoning, fast-moving volumes and varieties of data into actionable information and business insights.
• **IBM InfoSphere Data Explorer** provides federated discovery, search and navigation over a broad range of data sources to help organizations get started quickly with big data initiatives and gain more value from their information.
• **Robust IBM data warehouse software and integrated systems** help simplify and accelerate the delivery of insights derived from your data.
• **IBM PureData™ System for Analytics** is a high-performance, scalable, massively parallel system that enables clients to gain deep insight from their data and perform analytics on enormous data volumes.
• **IBM PureData System for Operational Analytics**—part of the IBM PureSystems™ family—is an expert integrated data system designed and optimized specifically for the demands of an operational analytics workload.

**For more information**
To learn more about how IBM solutions can help your organization capitalize on big data, please contact your IBM representative or IBM Business Partner, or visit: [ibm.com/bigdata](http://ibm.com/bigdata)