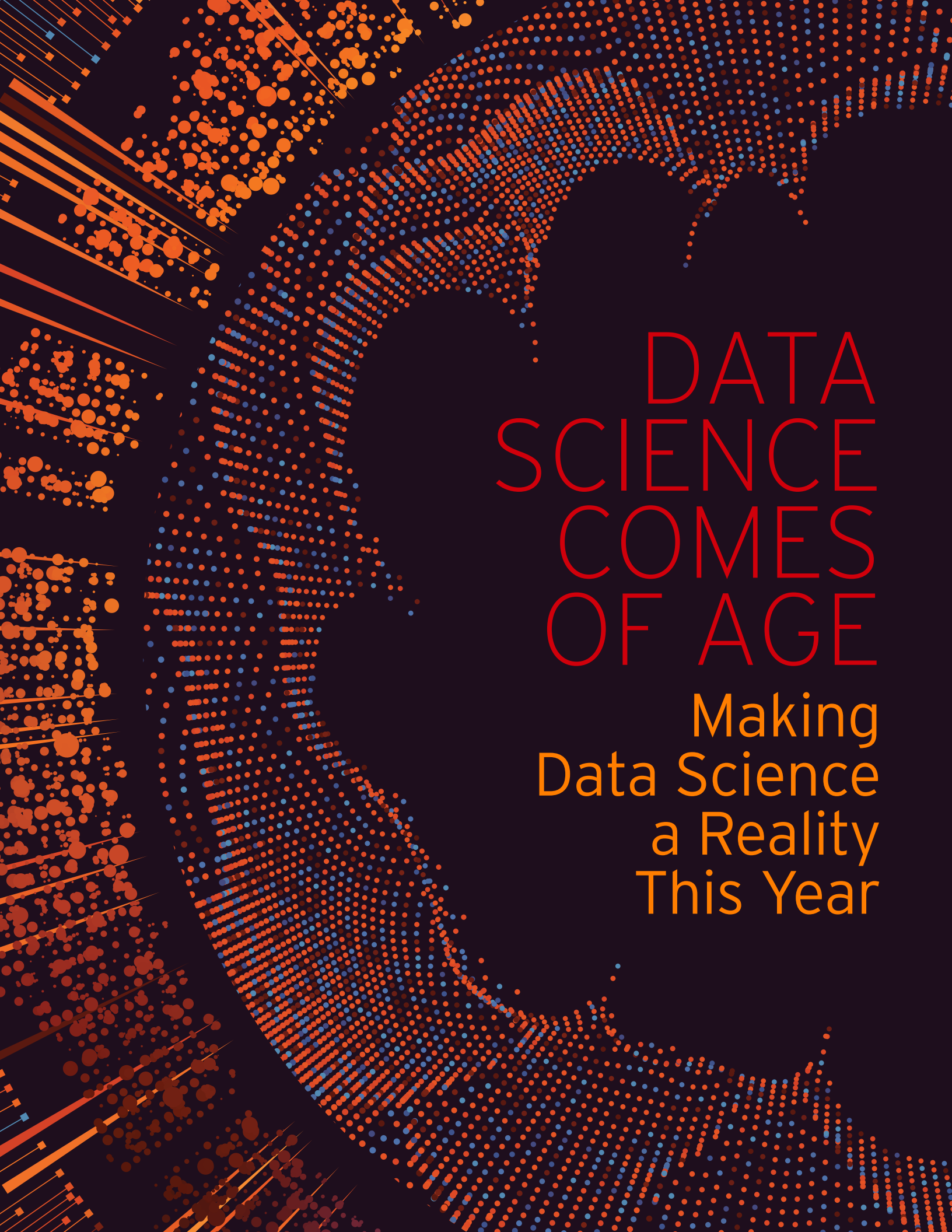
DATA SCIENCE COMES OF AGE
Making Data Science a Reality This Year

While data science has been on the radar of state and local government leaders for some time, the COVID-19 pandemic has highlighted the real-world implications and urgency of adopting artificial intelligence (AI), machine learning, robotic process automation (RPA) and similar tools that fall under the "data science" umbrella.

In a 2019 NASCIO survey, AI (in the form of machine learning, RPA and chatbots) was by far the top choice when respondents were asked which emerging technology would have the most impact in the next three to five years.[1] However, a variety of challenges prevented them from moving forward more fully with data science initiatives. Less than a year later, the pandemic created immediate demand for virtual assistants to help with skyrocketing call center volume, as well as geospatial mapping and predictive analytics to help manage COVID-19 outbreaks and other crises. In NASCIO's 2020 survey, 76% of respondents said they first introduced chatbots (virtual agents) in response to the pandemic.[2]

While the pandemic has jump-started data science programs, state and local governments still face roadblocks when rolling out and managing the technology and processes required to get the most value from their data troves. The good news is that organizations no longer need to delay plans to make data science a viable reality. Cloud technology, purpose-built hardware and pre-packaged data science software can get lone practitioners, data science teams and remote workers up and running within weeks if not days. Using these tools to rapidly make meaning of data, organizations can efficiently deliver personalized constituent services, offload work to digital twins, support planning and decision-making, and more.

Working Smarter While Wading in Data

The following challenges drive the need for innovative data science solutions.

☑ **Increasingly large data sets.** With the proliferation of Internet of Things (IoT) sensors, video imaging, smart phones and digital records, data sets are becoming larger and more complex. To extract full value from data, state and local governments must be able to quickly gather and apply insights. However, 42% of state CIOs say their data is not properly organized for use in AI or machine learning.[3]

Urgent demand for rapidly delivered, high-quality constituent services. The pandemic has forced the public to rely more heavily on digital government services. To deliver personalized services via call centers and web portals, organizations need to incorporate AI and machine learning into everything from chatbots and natural language processing to intelligently automated workflows.

While the pandemic has jump-started data science programs, state and local governments still face roadblocks when rolling out and managing the technology and processes required to get the most value from their data troves.

☑ **Doing more with less.** Back-office and customer-facing staff must provide increasingly complex services to a growing number of people. At the same time, budget cuts have forced some agencies to furlough workers or freeze hiring, and a wealth of process knowledge and subject matter expertise is being lost to a wave of retirements.

☑ **Talent shortages.** Skilled data scientists are expensive and hard to find. In traditional data science environments, a large portion of their time is spent on manual tasks such as collecting and normalizing disparate data rather than higher-order tasks such as modeling, analyzing and acting on data. To optimize practitioners' time, organizations need tools that streamline and automate processes wherever possible.

☑ **Infrastructure concerns**. Having sufficient bandwidth and processing power to perform data science operations is critical. Yet remote workers often rely on traditional laptops and other devices that have fewer built-in resources than onsite mainframes and workstations. In addition, a lot of valuable data resides in network or third-party silos that are difficult to access and integrate into data science systems.

☑ **Complexity of setting up and maintaining tools.** The data science toolkit contains a variety of tools for performing computations, manipulating data, creating and sharing visualizations, and more. Getting different versions of different tools to work together properly and securely delays implementation of data science solutions.

Unlocking the Power of Data from Day One

The following tools and strategies will help organizations roll out a robust data science program that enables organizations to agilely support a wide range of use cases at scale.

GEOSPATIAL MAPPING AND VIDEO IMAGES SOLVE REAL-WORLD PROBLEMS

Real-world examples of data science in state and local government are emerging in virtually every type of agency. One of the most common use cases for AI revolves around leveraging geospatial data for predictive analytics and real-time resource allocation. COVID-19 testing, tracking and vaccine scheduling rely on this approach.

Another application is in city planning and asset management. For example, city planners can analyze geospatial data about flooding or other weather-related events to zone housing and roadways in safer areas. They can also use that data to predict what resources (e.g., heavy equipment or public safety personnel) they'll need — and when and where — to respond effectively to those events. As they build out their vision, they can incorporate GIS-enabled IoT sensors and video cameras for real-time monitoring and decision-making related to things like river levels or potential bridge and roadway failures. Then, they can integrate that data with customer relationship management (CRM) applications to automatically alert key personnel, intelligently orchestrate response activities and deploy resources as situations evolve during an actual event.

Preloaded Data Science Software Stack

Data science practitioners can't start work until their computers are set up. In many cases, assembling, configuring, deploying and testing all the required software tools on a system can consume days if not weeks. In addition, siloed products may not integrate or provide the easy connectivity to cloud and other data resources that data scientists need to perform their jobs. Without in-person access to IT staff, remote practitioners may have to do more of their own IT problem-solving.

Preloaded data science software stacks are an increasingly popular option for overcoming these complexities. Loaded onto a cloud-connected laptop or a workstation designed for high-performance operations, these solutions come with all the needed tools in a typical data science workflow — from data preparation to model training and visualization. A command-line interface provides interoperability across hybrid cloud environments, allowing data scientists to access data coming from IoT sensors, cameras and cloud-based applications as well as back-office systems of record and other sources. By saving time on configuration, testing and troubleshooting, they allow staff to get to work right away.

Using Hybrid Cloud Environments Strategically

The two main systems for running software in a data science environment are a cloud-connected laptop or a workstation designed for high-performance operations. Hybrid cloud environments offer the best of both worlds, providing flexibility, scalability and seamless integration with disparate, widely distributed data sources. Workstations and laptops can work hand-in-hand to maximize productivity, accelerate time to insight and optimize costs.

Workstations — where data lives right on the machine — are ideal for scenarios where practitioners need to frequently interact with billions of rows of data and require extremely high-performance processing power. They allow users to efficiently prepare, clean and explore data and perform complex visualizations, simulations and modeling on a secure local device.

A cloud-connected laptop makes sense for scenarios where a user needs to gather data from remote data stores and perform routine analysis on smaller datasets that don't regularly require high-speed processing. Using cloud computing, organizations can invest in more portable, lower-cost, lower-performance laptop models for work at home or in the field. When users perform more intensive tasks, they can still draw on the cloud for additional processing power or storage.

Getting Started: Building a Data Science Practice

The following steps help organizations build a successful data science function.

☑ **Assess your data.** This process is the foundation for everything else. Determine what data you have access to, where it resides and how useable it is (e.g., is it accurate and up-to-date; is it structured or unstructured). Identify relevant security and compliance policies that apply to the data. Finally, decide whether the data provides the information you need.

☑ **Determine what you want to achieve with the data.** What types of questions do you want to ask or what decisions do you want to make based on the data?

☑ **Develop and enforce policies for the ethical use of data and AI.** When adhered to, these policies ensure questions are asked in a responsible, non-biased way; help build trust among the community; and protect the organization as well as its investment of time

HYBRID SOLUTION ENABLES AUTHORITIES TO RAPIDLY IDENTIFY SUSPECTS

A use case involving a large sporting event demonstrates how a hybrid cloud environment can be used strategically. A busy municipal airport needed to prepare for a major sporting event that would attract thousands of visitors to the large metropolitan area it serves. To protect residents, visitors and well-known athletes, the airport wanted to implement facial recognition technology that would spot known terrorists or other high-risk individuals in real time. The airport had more than 3,500 high-definition cameras throughout the airport, and wanted to apply AI-powered facial recognition technology to the data coming from the cameras.

Transferring all that camera data to the cloud for real-time AI processing was impractical. It would have required fiber optic cabling with extraordinary capacity. In addition, by the time the data was processed and results were returned to security personnel at the event, suspicious characters would be long gone.

To overcome these challenges, the security team set up a field station within the airport that used dozens of workstations with preloaded data science software. The team used the cloud to integrate threat data from international, national and local law enforcement sources and to train and model the facial recognition technology to identify those threats. Then they deployed those capabilities to the airport workstations, which could run the technology against a massive volume of camera images to rapidly identify suspects.

and resources into a project. Policies should address questions such as: Is what we're doing ethical; are we asking questions in a way that doesn't steer answers to a particular conclusion; and does this work align with privacy laws, the Bill of Rights and social justice concerns?

☑ **Ensure data is secure and complies with data privacy regulations.** Protecting the public's private information as well as sensitive research becomes more challenging as data leaves controlled onsite environments to enable remote work. Advanced data science workstations and laptops have built-in deep-learning tools that are trained to look for suspicious or anomalous behaviors and issue alerts when they detect a potential threat. Organizations may also want to consider multifactor authentication to validate users' identity, as well as role-based access control to ensure authenticated users can only access the data they are authorized to access.

☑ **Learn from peers.** Interacting with peers is a great way to stay up to date, share best practices, and identify new use cases and opportunities for data science. Consider joining a data science community such as Kaggle to download tools for manipulating and modeling data and to learn how other organizations solve data science problems.

Navigating the Future Intelligently

Demand for services that incorporate AI, ML and other data science approaches is likely to grow as state and local government leaders, workers and constituents gain a glimpse into the full potential of these tools. When used strategically, a hybrid cloud model that leverages state-of-the-art, purpose-built workstations and laptops – along with a preloaded data science software stack – can help organizations evolve more quickly to embrace a mature data science program.

This piece was written and produced by the Center for Digital Government Content Studio, with information and input from HP.

Endnotes:

1. NASCIO. 2019 State CIO Survey. The Responsive State CIO: Connecting to the Customer. October 2019. https://www.nascio.org/resource-center/resources/the-2019-state-cio-survey/
2. NASCIO. 2020 State CIO Survey. The Agile State CIO: Leading in a Time of Uncertainty.
3. NASCIO. What State and Local Government Technology Officials Can Expect. January 2020. https://www.pti.org/documents/Events%202020/NASCIO_2020TechForecast.pdf

Produced by: CENTER FOR DIGITAL GOVERNMENT

The Center for Digital Government, a division of e.Republic, is a national research and advisory institute on information technology policies and best practices in state and local government. Through its diverse and dynamic programs and services, the Center provides public and private sector leaders with decision support, knowledge and opportunities to help them effectively incorporate new technologies in the 21st century. **www.centerdigitalgov.com.**

For: hp

HP creates new possibilities for technology to have a meaningful impact on people, businesses, governments and society. With the broadest technology portfolio spanning printing, personal systems, software, services and IT infrastructure, HP delivers solutions for customers' most complex challenges in every region of the world. More information about HP (NYSE: HPQ) is available at **www.hp.com**