

Investing in a Better Storage Environment:

Best Practices for the
Public Sector

Industry Perspective

 **Hitachi Data Systems**
Federal



EXECUTIVE SUMMARY

The public sector faces numerous and known challenges that government employees work to overcome every day. Constrained budgets, ever-growing citizen expectations, information technology infrastructure issues — the list of obstacles that government agencies must deal with seems to grow constantly.

However, one challenge that is less obvious but is as important as any of the other difficulties the public sector is facing is the massive growth of unstructured data — especially where and how to best store it so that it's accessible and usable and retains its value.

Government data generation and archiving are on the rise because of the rapid growth of useful data sources like mobile devices and applications, smart sensors and Internet of Things devices, cloud computing solutions, and citizen-facing solutions. As this digital data increases and becomes more complex, the secure storage of it becomes ever more important as well. After all, the data that government and citizens generate today must live forever — stored, backed up and governed long after its original use is over. It must be discoverable, searchable and accessible, independent of the application or media with which it was created. And it needs to forever be available anywhere, anytime so that agencies and constituents can act on it at a moment's notice, even if that moment comes 20 years from now.

That's why GovLoop has partnered with Hitachi Data Systems Federal to create this industry perspective. Public sector data storage needs are at a tipping point and the time to act is now.

In this industry perspective, we will:

- Speak with and gain insights from Brian Houston, Vice President of Engineering at Hitachi Data Systems Federal.
- Explore the current landscape of public sector storage needs.
- Review solutions and best practices for the secure storage of unstructured data.
- Help the public sector best understand how it can deal with this explosion of data.

With the current massive growth of data that the public sector is facing, now is the time to learn how to get the most out of storage investments and prepare the public sector storage environment to stay agile as demands on IT grow in scope, complexity and cost.

THE CURRENT LANDSCAPE OF PUBLIC SECTOR STORAGE NEEDS

Storage requirements in the next 10 years are projected to grow to more than 40 times what they currently are. Will your agency's existing approach of meeting these storage requirements be sustainable and affordable? Have you been thinking about how to organize this massive growth of unstructured data and how to keep it accessible for your constituents to help meet mission need?

"With the advancement of cloud and big data initiatives as well as social media, the amount of data that the various agencies are creating is climbing dramatically," said Brian Houston, Vice President of Engineering for Hitachi Data Systems Federal. "If you look at the growth, it's exponentially expanded year over year."

Houston is correct. Over time, the public sector's data and storage needs have changed dramatically. Ten years ago, almost all data was structured. Transactions were manually entered; supporting paper documents were produced separately and archived in traditional file cabinets.

Today, of course, much of the government's business transactions are electronic, and digital transactions primarily include unstructured data. Supporting documents are kept electronically, emailed and made available via the Internet and prove to be a challenge to keep online and archived in a useful and logical way.

To better understand this growing need, you must understand the difference between structured and unstructured data.

"Unstructured data is anything outside of a database," Houston explained. "So when you start talking about unstructured data, you're talking about sensor-level data from different sensors, from machine-level data, from content-level data, like pictures, to Excel

docs, Word docs — anything that's sent through email attachments. Anything that you would put into your home drive or home directory, or anything that you'd put into a file share, would be the kind of data that is growing exponentially."

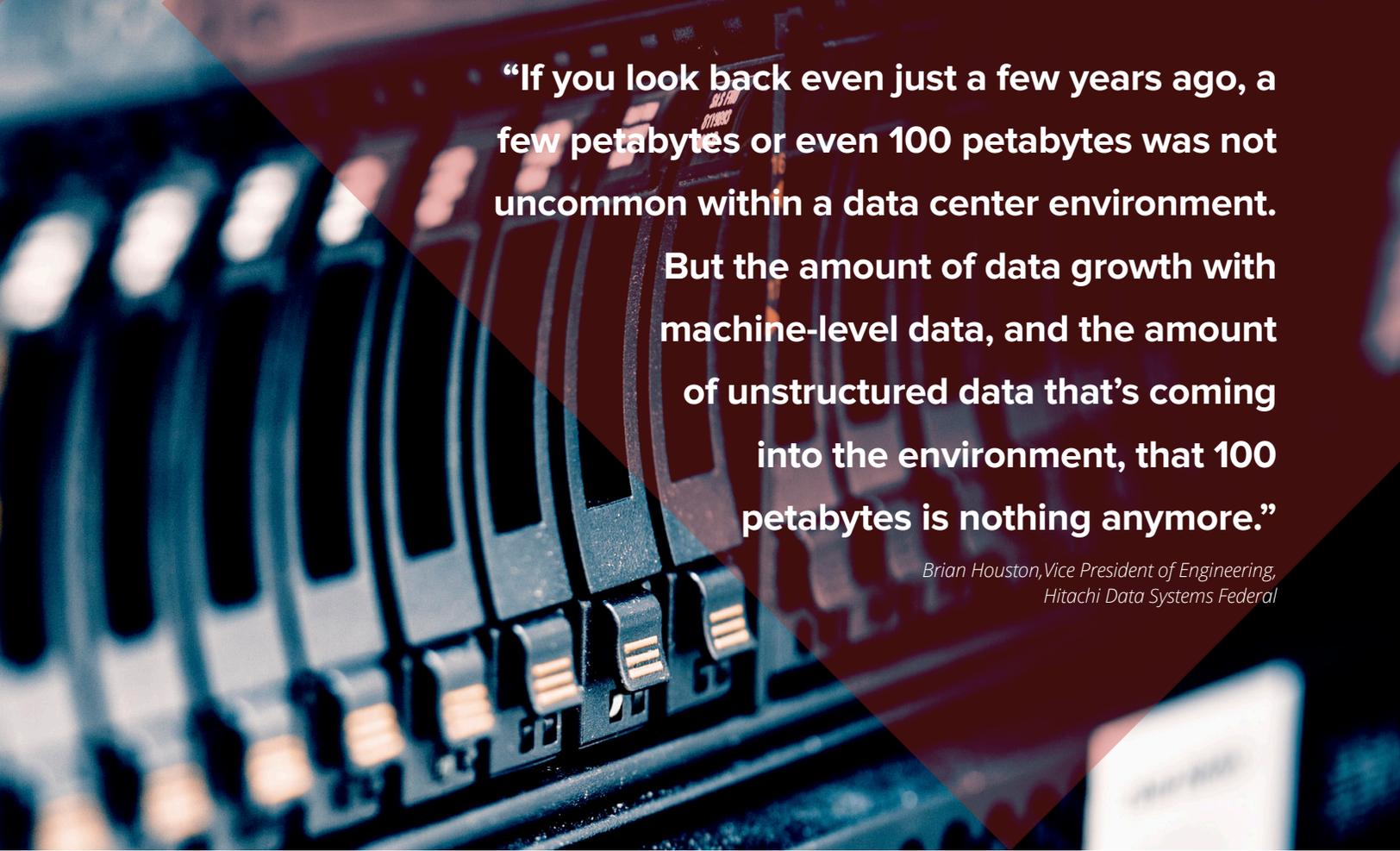
In short, data that does not fit well into relational tables such as databases is unstructured data.

For IT managers tasked with budgeting storage resources, the world has changed dramatically in a few short years. Along with storing structured data in databases, they must now also store, retrieve and protect mission-critical data that proliferates in file systems across the network.

This unstructured data is not just text and tables; it comes from a variety of applications. Along with "traditional" unstructured content, such as Microsoft Office files, Adobe PDFs and comma-separated value data, agencies are using new media content, including video, audio and graphics. Many public sector organizations use specialized applications that create a profusion of very large files. End users manage these applications, and the IT department does not have day-to-day visibility into their status.

"If you look back even just a few years ago, a few petabytes or even 100 petabytes was not uncommon within a data center environment," Houston said. But the amount of data growth with machine-level data, and the amount of unstructured data that's coming into the environment, that 100 petabytes is nothing anymore."

Today, he said, agencies are creating data that's growing to the exabyte level. Trying to understand where the data resides, what they have and how much of it is in duplication is key for agencies to save costs and keep up with the growth that is happening daily.



“If you look back even just a few years ago, a few petabytes or even 100 petabytes was not uncommon within a data center environment.

But the amount of data growth with machine-level data, and the amount of unstructured data that’s coming into the environment, that 100 petabytes is nothing anymore.”

*Brian Houston, Vice President of Engineering,
Hitachi Data Systems Federal*

“Think about it this way,” Houston said. “If you had 5 terabytes yesterday, and you have 100 users pumping in the same amount of data that they were doing on a daily basis, that 5 terabytes is going to turn into 10 or 15 terabytes, within a month. That exponential growth has to be contained, and you have to understand what you really have. It’s not just taking the data and putting it into a website or into a file share, and letting it keep growing.”

Now agencies must understand the data that’s coming in and ensure that it’s at the appropriate level of performance and availability. What’s more, they must make sure they know the age and lifecycle of the data to store it at the right tier and performance capability, he added.

To better illustrate this need for proper management and storage of unstructured data, Houston turned to a simple use case: email.

“You see an email and an attachment that comes in today, but a month from now are you even thinking about or looking at that email?” Houston asked. “That’s where the transformation of where does that data really need to reside and at what tier and at what level comes in.”

As Houston noted, an email you receive today is probably very important to you, but after about a week, it may become obsolete. Yet it must be retained in a discoverable fashion to maintain compliance.

“So that email should be moved down from a high-performance, high-scalability type of environment to a lower-tier disk in the lower-tier environment or even off into an optical platform where it’s still online, still available, still searchable,” explained Houston. “That will put it at the appropriate storage level.”

The consequences if this unstructured data is not stored properly can be serious for the public sector. Unstructured data grows at an incredible rate of 62 percent year over year, and Hitachi Data Systems Federal estimates that 93 percent of most organizations’ data will be unstructured by 2022.

If left unmanaged, the sheer volume of unstructured data that an agency generates each year can be costly in terms of storage. Additionally, the information contained in unstructured data is not always easy to locate. Some studies show that 25 percent to 35 percent of an employee’s time is spent looking for information and about half the time workers can’t find the data they need. This means that they either have to recreate it or deal with the consequences of missing critical information. The key is how to leverage this data to control access while sharing it among the individuals and teams who need it for collaboration.

When addressed with a proper storage and content platform, the possibilities for accessing and understanding the information contained within unstructured data are limitless.

ONE SOLUTION: THE HITACHI CONTENT PLATFORM

The Hitachi Content Platform (HCP) is a feature-rich, multipurpose content storage solution with the security, scale and broad application support to enable multiple workloads in one cluster. The platform provides the features and functionality to improve the delivery of a wide range of IT services.

By doing away with the complexity of mixed storage environments without sacrificing the ability to make use of third-party resources as part of overall storage approach, HCP helps customers get the most out of their storage investments and prepares the storage environment to stay agile as demands on IT grow in scope, complexity and cost.

Here's how the platform works, Houston said:

"Think of it this way: Say you have sensor-level data that is being collected by various aircraft. And then you also have satellite imagery that's being collected on a day-to-day basis. And each one of those is its own individual silo of storage. And each has its own analytics around it, and it's got its own search."

If a user could take all the disparate silos of data and combine them into a single data lake, or data pool, and put them into the content platform, they would then be able to run analytics across all of it at one time, Houston said.

Previous to that data being stored in HCP, he explained, you'd have to look at imagery data and sensor-level data, and also understand what the user was seeing. Then, you'd have to take the time to go through each one of those search criteria and each system.

"But with the Hitachi Content Platform," Houston said, "now I can go and run one search algorithm and get instantaneous results across all, be able to correlate everything at one time, and understand all the disparate systems at the same time, what that really means to the end user. That's really the type of solutions that we're bringing to the table."

HCP is a multipurpose, distributed, object-based storage system designed to support large-scale repositories of unstructured data. HCP enables IT organizations and cloud service providers to store, protect, preserve and retrieve unstructured content with a single storage platform. It supports multiple levels of service and readily evolves with technology and scales with needs.

With a vast array of data protection and content preservation technologies, the system can significantly reduce or even eliminate tape-based backups of itself or of edge devices connected to the platform. HCP obviates the need for a siloed approach to storing unstructured content. Massive scale, multiple storage tiers, Hitachi reliability, nondisruptive hardware and software updates, multitenancy, and configurable attributes for each tenant allow the platform to support a range of applications on a single physical HCP instance.

THE VALUE OF STORING DATA IN THE HITACHI CONTENT PLATFORM



AVAILABILITY

Get to data immediately without involving others.



RANDOM

Access data without having to run through a pile of irrelevant data.



AUTHENTICITY

Proof authenticity of electronic records.



SECURITY

Protect against unauthorized data access, data tampering and data disclosure when lost or stolen.



DISCOVERY

Search all data in free-form criteria in a timely fashion.

CONCLUSION

The facts are clear: Unstructured data will continue to grow at a high rate, which means IT administrators must find cost-effective, competitive ways to provide the data capacity agencies need while ensuring secure and consistent backup of mission-critical information. It's time to stop treating all data as equal, because it's not. The public sector must address its data storage needs now, as unstructured data continues to explode. Its relevance to the needs of government will only grow.

Technology now offers an exciting opportunity to transform the vast amounts of data that governments already collect into a deeper understanding of citizens' needs — and into forward-thinking, cost-effective ways to meet those needs. But for these analyses to be made, for mission need to be met and for citizens' needs to be understood, the first step is securely storing this data. Advancing in storage and management now will allow governments to glean more insights and relationships concealed within their data in the long term.

ABOUT AVNET

Avnet Government Solutions makes it easier and more affordable to enter and excel in the public sector markets locally and around the world. Let our Public sector and technology experts, strategic alliances, training, resources and services help you provide complete customer solutions that span the enterprise, data center and IT lifecycle. AGS has global reach and local expertise—offering partners and suppliers the best of both.



ABOUT GOVLOOP

GovLoop's mission is to "connect government to improve government." We aim to inspire public sector professionals by serving as the knowledge network for government. GovLoop connects more than 200,000 members, fostering cross-government collaboration, solving common problems and advancing government careers. GovLoop is headquartered in Washington, D.C., with a team of dedicated professionals who share a commitment to connect and improve government.

For more information about this report, please email us at: info@govloop.com



ABOUT HITACHI

Hitachi Data Systems Federal provides technology solutions that enable government agencies to extend the usable life of their IT infrastructure. By engineering technologies from the ground up, HDS Federal offers agencies greater reliability and scalability, while reducing total cost of ownership in budget-conscious environments.

Our industry leading virtualization and storage solutions enable 100 percent data availability wherever and whenever agencies need it. Whether building the pathway to the cloud or consolidating data centers, HDS Federal provides our customers with one platform for all missions in an age of disparate IT systems.

At HDS Federal, we implement data solutions that not only meet the needs of the government today, but more importantly, the government of tomorrow.

Hitachi Data Systems Federal



1152 15th St NW, Suite 800
Washington, DC 20005

Phone: (202) 407-7421 Fax: (202) 407-7501

www.govloop.com | [@GovLoop](https://twitter.com/GovLoop)