Big Data Express Data.gov

Issue No. 2, January 4, 2013

***Editor’s Notes:***

Happy New Year! 2012 was a big year for big data. From the White House to our homes, we are facing more data today than ever before. In this issue, I will share with you a new definition of big data, some suggestions to put on your to-do list for big data, and a free workshop on cloud and big data offered by NIST this month. Enjoy. –Jean Yan

**Recent Development: New Definition of Big Data**

(Source: [http://blogs.forrester.com/mike\_gualtieri/12-12-05-the\_pragmatic definition\_of\_big\_data](http://blogs.forrester.com/mike_gualtieri/12-12-05-the_pragmatic%20definition_of_big_data))

Current popular definitions usually use the “Vs” (volume, velocity, and variety, veracity, value, etc.) to describe big data. However, Mike Gualtieri, Principal Analyst at Forrester Research, considers these definitions “quaint” because they are only measurements. Such a definition is neither actionable nor complete for business and IT professionals. Gualtieri writes that exponential data growth makes it continuously difficult to manage — store, process, and access, that data contains nonobvious information that firms can discover to improve business outcomes. Further, measures of [data are relative](http://blogs.forrester.com/mike_gualtieri/12-12-05-is_750mb_big_data); one firm’s big data is another firm’s peanut. Gualtieri proposes a new “pragmatic” definition that looks beyond measurements alone:

Big Data is the frontier of a firm’s ability to store, process, and access (SPA) all the data it needs to operate effectively, make decisions, reduce risks, and serve customers.

An easy way to remember Gualtieri’s definition of big data is to think simply SPA:

* **Store.**Can you capture and store the data?
* **Process.** Can you cleanse, enrich, and analyze the data?
* **Access.** Can you retrieve, search, integrate, and visualize the data? V:\Documents and Settings\jeanyan.ENT.002\Local Settings\Temporary Internet Files\Content.IE5\W8AE81RP\MC900442171[1].png

# Feds in Action: NIST Joint Cloud and Big Data Workshop

**January 15-17, 2013**

(Source: <http://www.nist.gov/itl/cloud/cloudbdworkshop.cfm>)

### Registration: online, free (same website as Source above)

The NIST Cloud and Big Data Workshop will bring together leaders and innovators from industry, academia and government in an interactive format that combines keynote presentations, panel discussions, interactive breakout sessions and open discussion. The conference will be kicked off by Steven VanRoekel, the Chief Information Officer of the United States, and Pat Gallagher, Under Secretary of Commerce for Standards and Technology and Director, NIST.

The second and third days of the workshop will focus on the intersection of cloud and big data. Fully realizing the power of big data depends on meeting the unprecedented demands on storage, integration, and analysis presented by massive datasets–demands that cloud innovators are working to meet today. The workshop will explore possibilities for harmonizing cloud and big data measurement, benchmarking and standards in ways that bring the power of these two approaches to bear on driving progress and prosperity. V:\Documents and Settings\jeanyan.ENT.002\Local Settings\Temporary Internet Files\Content.IE5\W8AE81RP\MC900442171[1].png

## Forecast: To-Dos for Big Data in 2013

(Source: [http://www.datanami.com/datanami/2012-12-17/ 7\_things\_you\_need\_to\_do\_about\_big\_data\_in\_2013.html](http://www.datanami.com/datanami/2012-12-17/%207_things_you_need_to_do_about_big_data_in_2013.html))

Hardly a day goes by without the news mentioning big data – so what should we expect to do to succeed with big data in 2013? Here are some suggestions:

**1. Visualize**

Explore data visually. Data visualizations tell stories that spreadsheets simply can’t. The cycle of visual analysis is an interactive process of getting data, structuring it one way, noticing results and asking follow-up questions. These follow-up questions might lead to a need to drill down, drill up, filter, bring in new data, or create another view before you share and act on the data. You can do two things at any moment: (a) instantly change what data you are looking at, and (b) instantly change the way you are looking at it. This combination creates an exploratory experience required for anyone to answer questions quickly.

**2. Simplify**

While we already find it difficult to keep track of all the new database vendors and open-source projects, the landscape in 2013 will be even more crowded. To meet the challenge, try the following:

* Take a strategic approach by extending your relational and online transaction processing systems to one or more of the new on-premise, hosted or service-based database options that best reflect the needs of your industry/organization.
* Pick a real-time business intelligence platform that supports direct connections to over 30 database and file formats to enable you to pick the right tool for the job, so to accommodate the trend of increasing data volume..
* Choose the best mix for every project between connecting live to fast databases and importing data extracts into an in-memory analytics engine to offset the performance of slow or overburdened databases.

**3. Coexist**

Use the strengths of each database platform and enable them to coexist in your organization’s data architecture. Cloudera and Teradata jointly published a guide outlining requirements that are best suited for either a data warehouse or Hadoop.

**4. Empower**

Big data and self-service business intelligence go hand in hand, so empower everyone in your organization to use and benefit from big data. Make sure people can access and interact with data in dashboards right in their browser, or on a mobile device.According to Aberdeen Group’s recently published report, [***Go Big or Go Home? Maximizing the Value of Analytics and Big Data***](http://www.tableausoftware.com/learn/whitepapers/maximizing-value-analytics-and-big-data), “Organizations with big data are over 70% more likely than other organizations to have BI projects that are driven primarily by the business community, not by the IT group”.

**5. Integrate**

Consider how to integrate and blend data from disparate sources. Organizations that can blend different relational, semi-structured and raw data sources in real time without expensive up-front integration costs will be the ones that get value from their data. In an integrated world, spreadsheets, data warehouses and databases can all “talk” to each other. To ask and answer questions, your team members no longer need to know how data is stored, you have the flexibility to quickly connect to the data you need and consolidate it.

**6. Govern**

More than 80 countries have data privacy laws. The right balance between control and experimentation will vary depending on organization and industry. Authors from MIT-Sloan Management Review provide three tips for master data management in their article, [*Finding Value in the Information Explosion*](http://sloanreview.mit.edu/the-magazine/2012-summer/53409/finding-value-in-the-information-explosion/): (1) identify your “sacred data”; (2) define the workflows that will use unstructured data; and (3) use data to redefine business processes following a highly iterative process.

**7. Evangelize**

With the backing of one or more executive sponsors, evangelists get the ball rolling and instill a virtuous cycle – the more departments in an organization that realize actionable benefits, the more pervasive analytics becomes across the organization. Fast, easy-to-use visual analytics is the key that opens the door to organization-wide analytics adoption and collaboration. V:\Documents and Settings\jeanyan.ENT.002\Local Settings\Temporary Internet Files\Content.IE5\W8AE81RP\MC900442171[1].png

**End Note:** I would like to thank Sally Bourrie and Hyon Kim for proofreading this issue of *Express* and for their thoughtful comments. – Jean

V:\Documents and Settings\jeanyan\Local Settings\Temporary Internet Files\Content.IE5\8TVEN7YY\MC900059856[1].wmf