

SPONSORED  
BY:

MOBILE



Transform Your Business With Data

[ABOUT US](#)
[REGISTER](#)
[LOGIN](#)
[HOME](#)
[BLOGGERS](#)
[MESSAGES](#)
[POLLS](#)
[WEBINARS](#)
[RESOURCES](#)
[VIDEOS](#)
[STRATEGY](#)
[ANALYTICS](#)
[TECHNOLOGY](#)
[FB](#)
[TW](#)
[LI](#)
[G+](#)
[RSS](#)
[FINANCIAL SERVICES](#)
[HEALTHCARE](#)
[GOVERNMENT](#)
[EDUCATION](#)
[RETAIL](#)

## BIG-DATA PLATFORMS & TOOLS

### Islands in the Stream of Extreme Data



**Robert Plant**, Associate Professor, School of Business Administration, University of Miami  
Miami  
1/9/2013  
Comment  
7 comments

[Login](#)


50% 50%

The future of data is not just what is commonly termed "big data," but increasingly will be based upon real-time data streams. The challenge for CXOs and managers is going to be centered on understanding how to capture, manage, and drive value from these data streams.



Original uploaded by Pais (transferred by Frysch)  
(Source: [Wikimedia Commons](#))

As a society, we are all generating vast quantities of data. For example, our car signals how well we drive to our insurance company, how the engine has performed is recorded for use by the manufacturer, and an RFID pays our fees when we enter a London congestion zone. But multiply this by several million vehicles and their sensors and soon the data can get very "big," or "extreme" in nature.

#### Managing the stream

Streaming data is now becoming so large that companies need to understand, manage, and integrate it into their big data ecosystem. A Boeing airliner flying across the United States, for example, generates 240Tb of data on its journey, which, for a

company such as American Airlines, with a fleet of 600, could easily pass 144 Petabytes of data a day. The airplane is literally a data cloud in the clouds. Data can be relayed to the TSA for security; the FAA for flight information; the airlines MRO at each end and vendors can be directly alerted to ensure a part is available upon arrival if needed; customer systems such as arrival apps can be updated; and other supply chain partners linked in, including caterers, fuel, and cleaners.

#### Immense streams

Data streams have the potential to be immense. "System S" developed at the [Swedish Institute of Space Physics](#) to look at "Space Weather" generates 6 Gigabytes per second or 21,600Gb per hour -- equivalent to all the webpages of the Internet. While this may seem extreme, other companies are moving rapidly to mega if not yet extreme data status. Facebook's 600 million users spend 9.3 billion hours a month on the site and the average user uploads 90 pieces of data. At Wal-Mart, 1 million customer transactions are taken an hour. While Facebook and Wal-Mart's data is currently regarded as more transactional in nature, the volumes are more akin to streaming real-time data. This is one of the data transformational issues facing leading companies.

Amongst the primary drivers of streaming data are the technologies surrounding mobile platforms. Smartphone ownership has reached approximately 4.6 billion, each of which creates their own data stream, for example, spatial location data. All of this data has high potential value, either for making more sales for the company, as an asset that can be sold, or as a data-as-a-service product offering.

#### Tapping the streams, not collecting big data

Organizations must therefore not get fixated that big data is static, nor solely originating as manageable data "feeds" from their traditional systems and social media in predictable ways. Rather, they must start to think that they will be tapping extra-high volume data "streams," possibly beyond the company's internal systems capabilities and even that of many third-party vendors.

Storing data at a rate of 21,600Gb per hour, as experienced at System S, thus becomes a question of how to process that volume to capture the value without missing the nuggets of gold.

Two approaches are possible. One is to ramp up the datacenter, acquire big pipes to feed the data into the server clusters, and manage the scale that way. This is nearly impossible to do, as it is like damming a river to look for gold, at first it's OK, results can be produced, but eventually that dam produces a lake so big and deep that the gold is hard to find. Looking for patterns in extreme data sets by matching incoming data to the existing set in real-time is so computationally intense as to be practically impossible.

#### MORE BLOGS FROM ROBERT PLANT

##### A Review: Big Data in 2013

39 comments

Data scientists, NSA, C-suite buy in... Big data changed a lot in the last 12 months.

##### What Would Seymour Cray Think of Big Data?

13 comments

How would the tech legend respond to the 4 Vs?

##### The 360 Degree View of a Customer

56 comments

Big data is helping companies like Netflix see what you really want (and even when you go to sleep).

##### Lessons for Big Data From President Obama's Healthcare Implementation

24 comments

Three big data lessons for CEOs from HealthCare.gov's bumpy start.

##### Is Big Data Ruining Sports?

29 comments

From sailing to the NBA -- big data could increase the gap between sporting haves and have-nots (while ruining the magic in the process).

[More from Robert Plant](#)


#### FLASH POLL

[All Polls](#)

#### DIGITAL AUDIO



#### Big-Data Platforms & Tools Radio

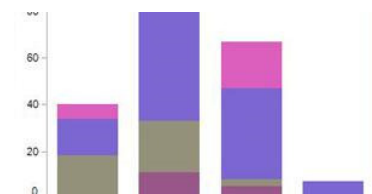
Latest Archived Broadcast  
[Who Can See From Your Platform?](#)

We discuss the best internal setups to make sure the right sets of eyes are watching the data development.  
[Listen](#)

[Full schedule](#)
[Archived shows](#)
[Download to calendar](#)

#### DATA VISUALIZATION SHOWCASE

##### Visualization: What's Stopping Big Data Improvement?



Survey results show bigger companies still confident in spite of massive IT overhaul requirements. Explore this data [here](#).

[More Data Visualization Showcase](#)

#### BDR IN YOUR INBOX

A second approach is to use data contextualization, a method developed by Jeff Jonas and data scientists at IBM. In this approach, incoming data is matched through Meta knowledge to the existing data set. While this is difficult at first, as the data sets are sparse and the connections therefore weak, as data is added, this task becomes easier as there is more contextual knowledge in the data set and the matching process actually speeds up rather than slowing down. This is a rather counterintuitive outcome and one that will allow the pioneers in the extreme data space to gain a competitive advantage over their rivals who build data dams to search for value.

It is worthwhile as we enter 2013 for CIOs and managers to start to consider the data streams within their firms that can be tapped, and how much data they will produce at what velocity. Examples include: trading systems, stock quote system, telemetry, and RFID data. In fact, any device connected to the Internet of Things can potentially be the source of a data stream. The second step is to consider how all these data streams connect, and what their contextual meta-data is that can be used to link them together. The final step is to start to build business models that this data stream could, if tapped, start to provide insights on, e.g., a parcel delivery routing problem, load balancing on a network for a telephone company, or real-time inventory management.

Once the company starts to think of its data within the context and the problems that can be solved through collections of contextualized streaming data, then a whole new dimension to the big data landscape can be opened up and processes old and new leveraged to profitable effect. Early movers are already experimenting with the technology, and so in order to keep up, this may be the time to put this topic high on your company's "to-do" list for the first quarter of 2013.

**Related posts:**


- [Welcome to the World of Openness & Big Data](#)
- [Stop Wasting Time on Visualization](#)
- [Big Sources: Sensor Networks Will Make Data Bigger & Cheaper](#)

— Robert Plant, Associate Professor, *School of Business Administration, University of Miami*

[Email This](#) [Print](#) [Comment](#)



**COMMENTS**

[Newest First](#) | [Oldest First](#) | [Threaded View](#)

 Keith Grinstead, User Rank: Petabyte Pathfinder  
1/30/2013 | 9:07:27 PM

**Re: You're scaring me...**  
I am not sure whether this should be islands or oceans of data?


[Login](#)

   
50% 50%

At least with oceans, assuming a bottomless ocean floor, we could imagine the depth of data that could be held. With these extreme streams (I like the water analogy here!) of data where do you see these going?



Does the corporate have to have the storage capacity to handle them or do you see them heading off into the cloud (a good analogy for your airline example!)?

[Reply](#) | [Post Message](#) | [Edit/Delete](#) | [Messages List](#) | [Start a Board](#)


 Robert Plant, User Rank: Blogger  
1/10/2013 | 9:20:45 PM

**Re: You're scaring me...**  
Not to extend the analogy too far but what if the movie is one of those that has four or more point of views and you just sample one... could lead to sub prime outcomes. Big data sampling is not simple and places a greater overhead on the system, but can be the only effective feasible solution to extreme, endless data streams from instruments.

[Login](#)



   
50% 50%

[Reply](#) | [Post Message](#) | [Edit/Delete](#) | [Messages List](#) | [Start a Board](#)

 Saul Sherry, User Rank: Blogger  
1/10/2013 | 5:04:36 AM

**Re: You're scaring me...**  
Wouldn't this just turn to the whole hypothesize, test and evaluate approach? If it fails you need to examine more frames... of course there's a lot to be said for looking at enough frames to get it right first time.

[Login](#)



   
50% 50%

[Reply](#) | [Post Message](#) | [Edit/Delete](#) | [Messages List](#) | [Start a Board](#)

 Christian Prokopp, User Rank: Blogger  
1/9/2013 | 3:23:13 PM

**Re: You're scaring me...**  
Also to be considered in this case is simply sampling data or aggregating data early on. Especially when it comes to sensor data the numbers quickly grow huge but the actual information content is a completely different matter.

[Login](#)

   
50% 50%

[Reply](#) | [Post Message](#) | [Edit/Delete](#) | [Messages List](#) | [Start a Board](#)


 Robert Plant, User Rank: Blogger  
1/9/2013 | 1:50:06 PM

**Re: You're scaring me...**  
You bring up an interesting question: can you judge an outcome from a frame in a movie or do you need to watch the whole thing?

[Login](#)



   
50% 50%

[Reply](#) | [Post Message](#) | [Edit/Delete](#) | [Messages List](#) | [Start a Board](#)

 legalcio, User Rank: Exabyte Executive  
1/9/2013 | 10:57:00 AM

**Re: You're scaring me...**  
So, @Robert and @Saul, are we looking at two distinct Big Data uses? Business (and, presumably, the research sector, weather, science, etc.) can have a snapshot of information based on real time streaming and adjust strategy accordingly, together with deep historical information that can produce trend-related data.

[Login](#)

   
50% 50%

[Reply](#) | [Post Message](#) | [Edit/Delete](#) | [Messages List](#) | [Start a Board](#)

**FOLLOW US ON TWITTER**

 **Big Data Republic**@BigDataRepublic 15 Apr  
AllJoyn: A Common Language For Internet Of Things - Open-source lingua franca aims to let IoT devices and services... [ow.ly/2FHbY7](#)  
Retweeted by Yakamoto  
[Expand](#)

 **Hajime Ozaki**@hajozaki 22h  
BDR Top 20 #BigData David Smith David Smith List of top tweeters on Big Data from @BigDataRepublic [wp.me/p4Aavi-2](#)  
[Show Summary](#)

 **Marcus Borba**@marcusborba 28 Apr  
#BigData #Analytics #BusinessIntelligence #FS @informationweek @AllAnalytics @BigDataRepublic @BigDataNetwork @BigDataClub

**LIKE US ON FACEBOOK**


 **Big Data Republic**  
Like

4,168 people like Big Data Republic.



Facebook social plugin

**ACCOLADES**

 **MEDIA PIONEER AWARDS WINNER 2013**  
Media Pioneer Award Winner



Saul Sherry, User Rank: Blogger  
1/9/2013 | 10:42:58 AM

**You're scaring me...**

@Robert, we're only just settling into this idea of big data as an unweildy mess that we can actually weild, and now you tell us the future isn't static?? Looks like the revolution in how businesses regard their data will have to continue.

[Login](#)



50% 50%

[Reply](#) | [Post Message](#) | [Edit/Delete](#) | [Messages List](#) | [Start a Board](#)

**BIG DATA REPUBLIC**

[ABOUT US](#) | [CONTACT US](#) | [HELP](#) | [REGISTER](#)  
[TWITTER](#) | [FACEBOOK](#) | [LINKED IN](#) | [GOOGLE+](#) | [RSS](#)



bigdatarepublic : /bigdatarepublic/section/2635 : /bigdatarepublic/section/2635/256757



UBM TECH

OUR MARKETS: [Business Technology](#) | [Electronics](#) | [Game & App Development](#)

Working With Us: [Advertising Contacts](#) | [Event Calendar](#) | [Tech Marketing Solutions](#) | [Corporate Site](#) | [Contact Us](#) / [Feedback](#)

[Terms of Service](#) | [Privacy Statement](#) | [Copyright © 2014 UBM Tech, All rights reserved](#)