



HADOOP MANAGEMENT MADE EASY

Provision, Operate, Monitor
Manage, Scale with Ease!

FREE EBOOK

datanami

• BIG DATA • BIG ANALYTICS • BIG INSIGHTS •

(<http://www.datanami.com>)

[Home](http://www.datanami.com/) (<http://www.datanami.com/>) [About](http://www.datanami.com/about/) (<http://www.datanami.com/about/>)

[Whitepapers](http://www.datanami.com/whitepaper/) (<http://www.datanami.com/whitepaper/>) [Events](http://www.datanami.com/events/) (<http://www.datanami.com/events/>)

[Subscribe](http://www.datanami.com/subscribe/) (<http://www.datanami.com/subscribe/>)



Select Language ▾

Translation Disclaimer ([/termsofuse.html#translation](#))

Search this site

Search

Follow Datanami: [f](http://www.facebook.com/pages/Datanami/124760547631010)

(<http://www.facebook.com/pages/Datanami/124760547631010>)

[Twitter](http://www.twitter.com/datanami) (<http://www.twitter.com/datanami>) [LinkedIn](http://www.linkedin.com/groups/Big-Data-News-4166980) (<http://www.linkedin.com/groups/Big-Data-News-4166980>)

[RSS](http://www.datanami.com/feed/) (<http://www.datanami.com/feed/>)

HOME (<HTTP://WWW.DATANAMI.COM>)

FEATURES

SECTORS

APPLICATIONS

TECHNOLOGIES

VENDORS

JOB BANK (<HTTP://WWW.DATANAMI.COM/JOB-BANK>)



September 11, 2015

Unstructured Data Analytics Shouldn't Be Such a Mess

Kon Leong



(<http://www.enterprisetech.co>)



(<http://www.hpcwire.jp>)

Top News from
Leading Solution
Providers
(<http://www.datanami.com>)



The problem with searching for a needle in a haystack is that the process, by nature, is inefficient. So why has it become a popular analogy for analytics efforts within the enterprise? Because today's analytics attempts – particularly for unstructured human data – are typically a mess.

Today's analytics are often ad hoc and rely on incomplete or skewed sample sets. They commonly focus on only one narrowly-defined data type. So for each glimmering "needle" of insight, it seems that heaps of data are scattered and cast aside, often at the cost of subsequent business efficiency.

When it comes to business content such as files, email, social media, IMs, calendar entries, images, and more, firms are failing to extract meaningful insight despite the potential wealth of information contained within.

There's a reason for this: Because the data is poorly managed to begin with. Businesses are treating analytics as a separate business function from data governance, when it's actually fundamentally dependent on it. Analysis occurs downstream; so by neglecting initial information governance infrastructure and practices, the enterprise is essentially sampling tiny random buckets of data from a whitewater river of information.

Many firms struggle to manage or even understand what sort of unstructured content they even have, let alone begin to effectively manage it. History is partially to blame, to be sure; most attempts at managing unstructured content were hastily prompted by waves of regulatory and legal reform that demanded immediate action. A reactive response was triggered, and many of those initial "band-aid" information management fixes remain in place today. Simply scratching beneath the surface often reveals a tangled mess of siloed

data platforms such as enterprise content management systems (ECMs), legacy systems, duplicated copies, and even entire missing categories of data.

It's no wonder that businesses are having trouble getting analytics value from this data.

For unstructured data, it makes sense; the technology required to process large volumes of diverse human-generated content is nascent compared to traditional BI and ERP systems that mine more mature and structured forms of data. Add to that the general state of mismanagement of most unstructured content, and you have an environment in which data never reaches its potential ... or worse, results in erroneous business decisions.



(<http://2s7gj373w3x22jf92z99mgm5w.wpengine.netdna-cdn.com/wp-content/uploads/2015/09/haystacks.png>)

Don't hate the hay — hate having multiple stacks of hay

This Just In | Most Read

September 15, 2015

Metanautix Partners With Sugon
(<http://www.datanami.com>this-just-in/metanautix-partners-with-sugon/>)
Tableau 9.1 Now Available
(<http://www.datanami.com>this-just-in/tableau-9-1-now-available/>)



Taking Action with
Real-Time
Enterprise Analytics

Download Now

Sponsored Whitepapers

Teradata Aster, Machine Learning, and Big Data
(<http://www.datanami.com/whitepaper/teradata-aster-machine-learning-and-big-data/>)

SpotlightON: Accelerating Time to Value of Hadoop Data
(<http://www.datanami.com/whitepaper/spotlighton-accelerating-time-to-value-of-hadoop-data/>)

View the Whitepaper Library
(<http://www.datanami.com/whitepaper/>)

Sponsored Multimedia



We don't necessarily need to get rid of the extra data "hay" that these stacks are composed of; that would entail getting rid of potentially valuable content. We just need to completely re-think how the data itself is managed. No more data "haystacks" means no more disparate data sources, no more ad hoc sampling attempts, and no more dirty or duplicated data. Furthermore, it vastly reduces the compliance risk associated with data mismanagement; with consolidated control, policies for management and eventual disposal can be implemented centrally and securely.

The lesson here is that data governance is the necessary foundation of all successful data analysis. The statistics axiom of "garbage in, garbage out" is used ad nauseam for a reason: because it's accurate and timelessly relevant.

So at risk of abandoning our original metaphor, we're trying to build a data lake, pooling all available resources into a single environment where they can be managed and analyzed in real time. Forward-leaning businesses are already making strides to achieve this, and they're not doing it with flashy analytics tools – those can come later, once the foundation is built. In a cohesive governance environment, analytics can be brought TO the data, rather than data cumbosomely being sampled and brought TO the stand-alone tools.

If large organizations want better analytics, they need to start with better information governance practices. After all, they probably should have been better all along.



[About the author:](http://2s7gjr373w3x22jf92z99mgm5w.wpengine.netdna-cdn.com/wp-content/uploads/2015/09/Headshot-Kon-Leong.jpg)

Kon Leong is CEO and Founder of ZL Technologies

[\(http://www.zlti.com/\)](http://www.zlti.com/). For two decades, he has been immersed in large-scale information technologies to solve big data issues for enterprises. His focus for the last 14-plus years has been on massively scalable archiving technology to solve records management and eDiscovery challenges for the government and private sectors. He speaks frequently at records management and eDiscovery conferences on cutting edge trends and solutions. A serial entrepreneur, Mr. Leong earned a BS degree from Loyola (Concordia U) and an MBA from Wharton (U of Penn).

Related Items:

Beware the Dangers of Dark Data (<http://www.datanami.com/2015/08/18/beware-the-dangers-of-dark-data/>)

Pulling Insights from Unstructured Data – Nine Key Steps
(<http://www.datanami.com/2015/01/16/pulling-insights-unstructured-data-nine-key-steps/>)

Share this:

(<http://www.datanami.com/2015/09/11/unstructured-data-analytics-shouldnt-be-such-a-mess/?share=twitter&nb=1>)

(<http://www.datanami.com/2015/09/11/unstructured-data-analytics-shouldnt-be-such-a-mess/?share=facebook&nb=1>) 13

(<http://www.datanami.com/2015/09/11/unstructured-data-analytics-shouldnt-be-such-a-mess/?share=google-plus-1&nb=1>)

(<http://www.datanami.com/2015/09/11/unstructured-data-analytics-shouldnt-be-such-a-mess/?share=linkedin&nb=1>) 23

(<http://www.datanami.com/2015/09/11/unstructured-data-analytics-shouldnt-be-such-a-mess/?share=pocket&nb=1>)

(<http://www.datanami.com/2015/09/11/unstructured-data-analytics-shouldnt-be-such-a-mess/?share=reddit&nb=1>)

(<http://www.datanami.com/2015/09/11/unstructured-data-analytics-shouldnt-be-such-a-mess/?share=pinterest&nb=1>) 1

(<http://www.datanami.com/2015/09/11/unstructured-data-analytics-shouldnt-be-such-a-mess/?share=tumblr&nb=1>)

(<http://www.datanami.com/2015/09/11/unstructured-data-analytics-shouldnt-be-such-a-mess/?share=email&nb=1>)

Tags: analytics (<http://www.datanami.com/tag/analytics/>), big data (<http://www.datanami.com/tag/big-data/>), haystacks (<http://www.datanami.com/tag/haystacks/>),

Integrate the Art of Data Science into Your Business Community

Webcast Sponsored by: TERADATA

[\(http://www.datanami.com/the-art-of-data-science-into-your-business-community/\)](http://www.datanami.com/the-art-of-data-science-into-your-business-community/)

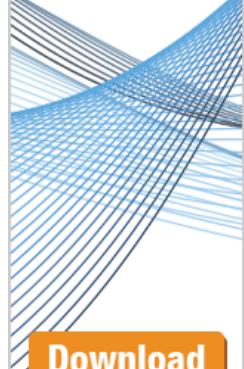
Integrate the Art of Data Science into Your Business Community

the-art-of-data-science-into-your-business-community/

Integrate the Art of Data Science into Your Business Community

the-art-of-data-science-into-your-business-community/

Streamline Manufacturing with Virtual Product Design



Download this Whitepaper



www.sgi.com Built with the latest Intel® Xeon® processors

Contributors

Alex Woodie
Editor in Chief

George Leopold
Contributing Editor

Steve Conway
IDC