Journal of Applied Economics and Business

OPEN ACCESS ARTICLE

How can Big Data transform knowledge management?

András Lévai*[1]

[1]RGDI, Széchenyi István University, Győr, Hungary

*info@tudasmenedzsment.hu

Abstract

This document examines the new technological hype, the big data by analyzing the road that led to big data. After discovering trends in storage costs and integrated circuit number of parts, is a description of the core theory regarding to big data, the 4V. Findings suggest that big data is usable at knowledge management, it is possible and useful to monitor collect data from our users to improve the knowledge management system use.

Keywords

Big Data; Knowledge Management, Social Intranet, ELGG

INTRODUCTION

On an usually day we are browsing, sharing, searching, communicating, buying on the Internet. These activities have a trace, creates large and complex data. These data is saved because the availability of cheap, fast computers and storage, as well as open source tools.

The road to big data

Two factors assisted reaching the road to big data. First, the processing power is very fast nowadays. Intel co-founder Gordon E. Moore described a trend that the number of transistors on integrated circuits doubles approximately every two years.

The cost of storage is almost null. Matthew Komorowski [9] decided to look for some historic pricing information to see exactly how fast the cost of storage space has gone down over the last 30 years. He used a web page called Historical Notes about the Cost of Hard Drive Storage Space [8]. For data from 2004-present he was retrieved using the archive.org Wayback Machine. Matthew did this research in 2009, so, in this research we enhance it. In this respect, we added 2010-2013 and we introduce a second chart to visualize the storage cost per terrabyte.

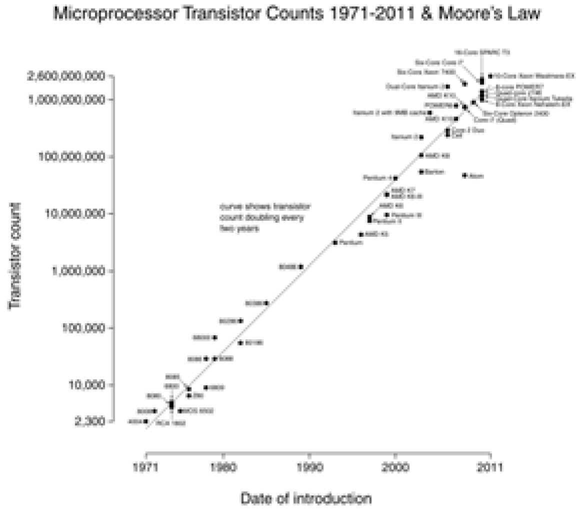


FIG. 1 MOORE'S LAW

In 2011 there was a flood in Thailand which created a shortage, so the prices increased for a year. Hard drive prices touch pre-flood levels in 2012 November [10].

So all these led to big data, now is available to use cheap hardware to store and process the information in innovative forms. The origins of the big data term come from a 2001 paper by Doug Laney of Meta Group [11]. In the paper, Laney defines big data as data sets where the three Vs—volume, velocity and variety—present specific challenges in managing these data sets. According to Ohlhorst [2] big data has 4V:

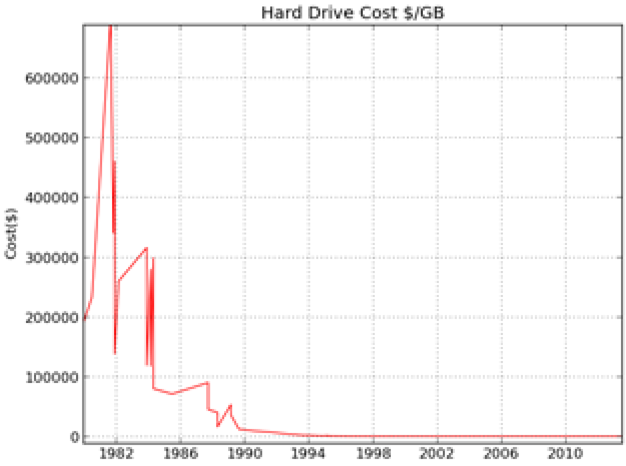


FIG. 2 HARD DRIVE COSTS PER GIGABYTE

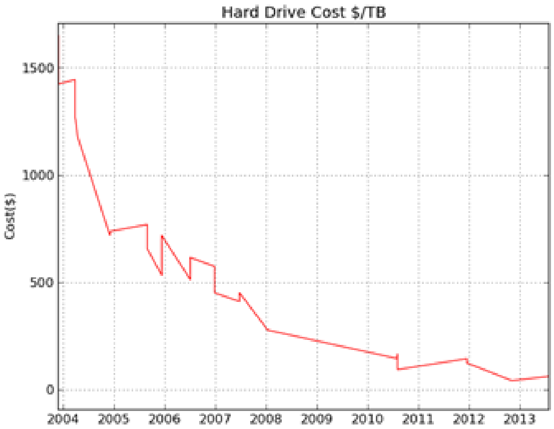


FIG. 3 HARD DRIVE COSTS PER TERRABYTE

1. Volume. Big Data comes in one size: large. Enterprises are awash with data, easily amassing terabytes and even petabytes of information.

2. Variety. Big Data extends beyond structured data to include unstructured data of all varieties: text, audio, video, click streams, log files, and more.

3. Veracity. The massive amounts of data collected for Big Data purposes can lead to statistical errors and misinterpretation of the collected information. Purity of the information is critical for value.

4. Velocity. Often time sensitive, Big Data must be used as it is streaming into the enterprise in order to maximize its value to the business, but it must also still be available from the archival sources as well.

Big data and data mining can be used to predict disease outbreaks, understand traffic patterns, and improve education. To show this we collected a list with real life examples of successful Big Data projects (see Table 1).

TABLE 1. BIG DATA EXAMPLES

Company	**Big Data Source**	**Findings**	**Url**
Ford	Fusion generates up to 25 GB data per hour	To understand driving behaviors	http://www.datanami.com/datanami/2013-03-16/how_ford_is_putting_hadoop_pedal_to_the_metal.html
Caesars casino		found that increasing pay within certain limits had no impact on turnover	Phil Simon: Too Big to Ignore [7]
Union Pacific Railroad	Ultrasound scanners sending every passing train and send the data to the railroads data center	identify equipment at risk of failure	http://en.wikipedia.org/wiki/Industrial_Internet
Walmart	Using big data from websites to feed shopper and transaction data into an analytical system	Shoppycat product	http://gigaom.com/2012/03/23/walmart-labs-is-building-big-data-tools-and-will-then-open-source-them/
Google	Search terms used	Flu map	http://www.google.org/flutrends/
Tesco		Cut cooling costs	http://www.computerweekly.com/news/2240184482/Tesco-uses-big-data-to-cut-cooling-costs-by-up-to-20m
Xerox		Xerox found that experience was overrated for call-center positions. What's more, overly inquisitive employees tended to leave soon after receiving training.	Phil Simon: Too Big to Ignore [7]
EMC Corporation	Insurance data	Better car insurance	Phil Simon: Too Big to Ignore
Thomas M. Menino (Boston's mayor)	Mobile phone gps data	Finding potholes and general road hazards	http://streetbump.org/
Amazon	Mining customer data	Recommender system	http://www.bigdata-startups.com/BigData-startup/amazon-leveraging-big-data/

Knowledge Management 2.0

The research topic is knowledge management 2.0, which is based on the foundations of social web. Tumblr, Twitter, Facebook, Instagram changed our life, we posting our thoughts, we share the photo immediately online right after we shot it. First it was a tool to communicate, and then an agent of change, we should think about the Egypt's Facebook revolution. It's not just a place to discuss about politics, it transforms people life, creates new industries. Flash mobbing show how efficiently is possible to organize an event. Web 2.0 provides new systems for knowledge management. People are now the generator of knowledge, they social participation grows, they can communication in a many-to-many model, the culture is changed so they are now more proactive, thanks to the possibility of self-organizing. Social network is also a tool to improve the transfer of implicit knowledge. According to Stowe Boys implicit knowledge is an interpersonal knowledge, which is communicated implicitly in the conversations and connections of people. Knowledge management 2.0 is focusing on socialization the most important mode of knowledge creation [1].

Széchenyi István University's Regional Development Doctoral School (RGDI)

RGDI has around 170 members, who can inform about new publication possibilities, call for papers, and conferences in a newsletter. This communication method makes hard to enhance the conversation on topics. So there was a need for a change. We chosed ELGG. "Elgg is an award-winning open source social networking engine that provides a robust framework on which to build all kinds of social environments, from a campus wide social network for your university, school or college or an internal collaborative platform for your organization through to a brand-building communications tool for your company and its clients." [3] Elgg is aimed primarily at education, according to the developer Dave: "Elgg focuses on the learner and interactions whereas VLE's focus on the course and content delivery. It's about providing an informal space that lets learners exercise their own thoughts, reflections, make their own connections and be able to compile a body of evidence that would normally slip through the cracks with the more highly structured approach that a VLE offers. The creation of ad-hoc communities around similar interests is what happens when you learn and discuss in real life, and Elgg allows people to do this in the online space, whereas Virtual Learning Environments do not." [4]

Sofar, we learned from the big data concept, that every log, database contains valuable data, so we examined how is it possible to make research about the ELGG data. In this line, we checked the data schema (see Table 2).

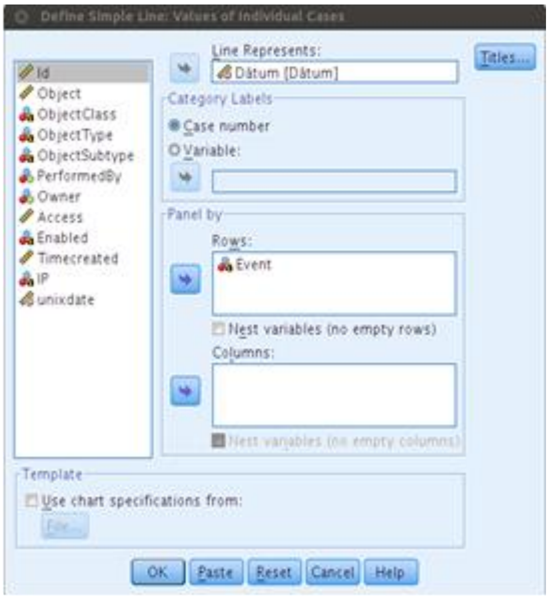
TABLE 2. ELGG DATA SCHEMA

Table name	Function
elgg_access_collection_membership	connection between access collection and user tables
elgg_access_collections	contains the user access
elgg_annotations	empty table
elgg_api_users	empty table
elgg_config	System parameter can be found here
elgg_datalists	path and cache information are stored here
elgg_entities	entity table
elgg_entity_relationships	shows the relation between entities
elgg_entity_subtypes	description of entities
elgg_geocode_cache	empty table
elgg_groups_entity	user groups access level
elgg_hmac_cache	empty table
elgg_metadata	entity and metadata connection table
elgg_metastrings	metadata description
elgg_objects_entity	objects description
elgg_private_settings	settings
elgg_river	river table
elgg_sites_entity	it contains the site name and url
elgg_system_log	log table
elgg_users_apisessions	empty table
elgg_users_entity	user list
elgg_users_sessions	empty table

Table 2 is important due to fact that the system log is stored. The default system log is stored in the prefix_system_log database table. It contains the following fields:

- id A unique numeric row ID
- object_id The GUID of the entity being acted upon
- object_class The class of the entity being acted upon (eg ElggObject)
- object_type The type of the entity being acted upon (eg object)
- object_subtype The subtype of the entity being acted upon (eg blog)
- event The event being logged (eg create or update)
- performed_by_guid The GUID of the acting entity (the user performing the action)
- owner_guid The GUID of the user which owns the entity being acted upon
- access_id The access restriction associated with this log entry
- time_created The UNIX epoch timestamp of the time the event took place

Analysis


Define Simple Line: Values of Individual Cases

Line Represents: Dätum [Dätum]

Category Labels: ● Case number ○ Variable:

Panel by — Rows: Event

□ Nest variables (no empty rows)

Columns:

□ Nest variables (no empty columns)

Template: □ Use chart specifications from:

FIG. 4 ELGG DATA IN SPSS

The following analysis is very easy after importing the log tables data into SPSS (see Fig. 4):

1. transactions and event types
2. users and event types
3. users and transactions
4. time and transactions

Conclusion

It is hard to decide wheather Big Data is a hype or not, but the technology allows us to collect, crawl almost every data. This gives us more research opportunity, like how the knowledge transfers in a social intranet. My next paper will be a case study about the RGDI Social Intranet project. Maybe it won't transforms the knowledge management, but the data can be used to fine tune the knowledge management system.

REFERENCES

[1] A. Gandih, L'enterprise sociale: Utiliser les applications Enterprise 2.0 pour déculper la productivité des travailleurs du savoir. Oracle White Paper, CA, USA., 2008.

[2] F. Ohlhorst, Big data analytics : turning big data into big money, ISBN 978-1-118-23904-9, Published by John Wiley & Sons, Inc., 2013.

[3] Elgg Foundation project – About ELGG, Accesed: September 10, 2013. http://elgg.org/about.php

[4] S. O'Hear: Elgg - social network software for education, August 10, 2006. Accesed: September 10, 2013. http://www.readwriteweb.com/archives/elgg.php

[5] E. Dumbill, Big Data Now: 2012 Edition, O'Reilly, 2012.

[6] V. Mayer-Schönberger and K. Cukier, Big Data: A Revolution That Will Transform How We Live, Work, and Think, Harcourt Publishing Company, 2013

[7] P. Simon, Too Big to Ignore: The Business Case for Big Data by Phil Simon, WILEY, 2013

[8] A. Shugart, Cost of Hard Drive Storage Space, Accessed: September 10, 2013. http://ns1758.ca/winch/winchest.html

[9] M. Komorowski, A History of Storage Cost, Accessed: September 10, 2013. http://www.mkomo.com/cost-per-gigabyte

[10] N. Randewich, Thai floods boost PC hard drive prices, October 28, 2011. Accessed: September 10, 2013. http://www.reuters.com/article/2011/10/28/us-thai-floods-drives-idUSTRE79R66220111028

[11] D. Laney, 3D Data Management, Application Delivery Strategies, Meta Group, 2001