

Big Data processing and Semantic technology a symbiosis



Material Subject to Creative Commons License:



A Taxonic whitepaper, by Colin Meerveld @Taxonic taxonic.com

This work is licensed under the Creative Commons Attribution-NonCommercial- NoDerivs 3.0 Unported License. You are free to copy, distribute and transmit the work, under the following conditions: you must attribute the work mentioning the author and Taxonic; you may not use this work for commercial purposes; you may not alter or transform this work.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

For questions regarding usage, please contact info@taxonic.com

Contents

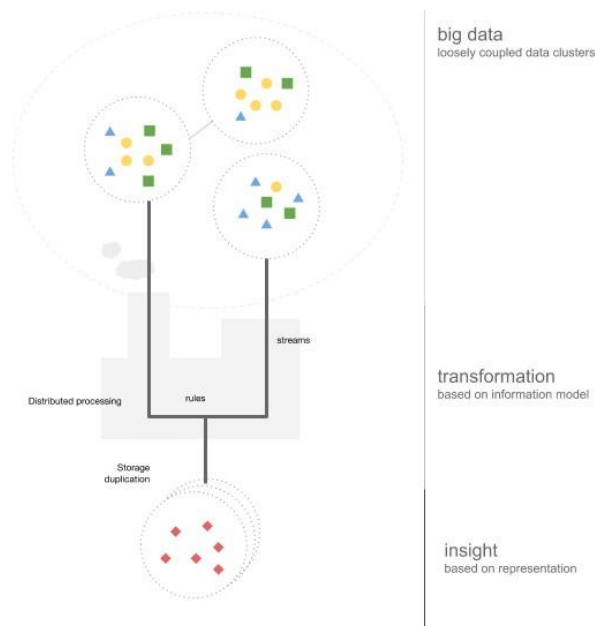
1	Big Data	4
2	Semantic technology.....	5
3	Complement traditional big data approaches with semantic technology.....	6
4	summary	7

1 Big Data

The core idea of big data is to collect data points and make sense out of it. Sufficient data point correlations may be found which helps analyzing causations.

However, when engineers talk about big data it quickly becomes a debate about software stacks—Hadoop versus Spark for instance. Sometimes it seems that there are no other options than to invest in new infrastructures for managing massive resource consumption: storage, bandwidth and computation to name a few. In this article I refer to this line of thought as the traditional big data approach.

As you may have guessed there are other approaches as-well. One of this is the semantic technology approach which I will explain in the next section. First, I provide a brief overview of the traditional approach.



The picture above depicts the transformation from big data to insight used for correlation and causation analysis and similar goals. Data is usually clustered, it may be exposed via a web service, logging system or by any other means. The clusters may be loosely coupled but more often stands on their own.

The first step is to translate the data into useful information—that is, adding meaning. In traditional big data approaches the data is translated according an information model. In this context, an information model is an informal way of defining the data. The data is translated using this model which may vary with each solution. Since not only the amount of data is big, but also the number of different sources, one has to deal with many data models. Creating the necessary transformation rules is an arduous task.

Translating large volumes of data is a challenging process. The data cannot be processed by a single computer. Also, since most data is duplicated, storage may be a bottleneck.

Furthermore, the rate of change may be high and will cause simple queries to underperform. In terms of big data, we often speak about data streams as the data may constantly change over time—for example in case of sensory data.

To cope with all those challenges, we often need specialized solutions which is far from trivial and lack of standardization.

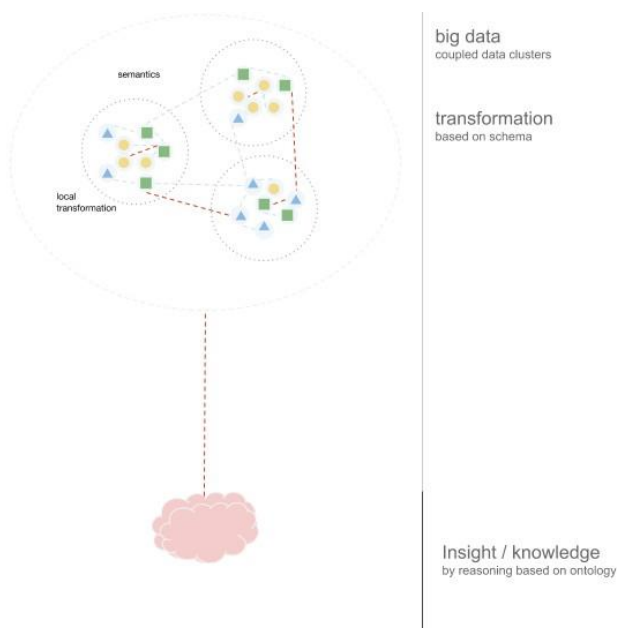
After translating the data into information, different representations could be generated to get insights. Note that those representations present only a limited view of the world.

2 Semantic technology

In contrast with traditional big data approaches, semantic technology will translate the data near to the source or (more useful) publishing it directly to the correct data structure. In any case the transformation is translated with domain knowledge and using international open standards. The all-important difference with the traditional approach is that the data models are machine-readable and available on the web. Consumers using the data for big-data analyses can therefore directly use the data, without the need to create and execute transformation rules.

Furthermore, the data is distributed by design: hence, the name semantic web. Processing and storage, however, is local, which makes resource consumption less of a pain. All the resource challenges you have with traditional approaches have vanished. In return you will get ... other challenges ;-)

Semantic technologies are based on relations between datapoint cross referenced with datapoints outside their data cluster. This is done by providing a unique identifier in the form of a URI to each datapoint. With this approach you create one unified data source. On top of the data source, you could reason and create so-called knowledge—a sort of machine insight based on logic. The image below depicts this idea. The dotted lines represent relations where the red ones are inferred.



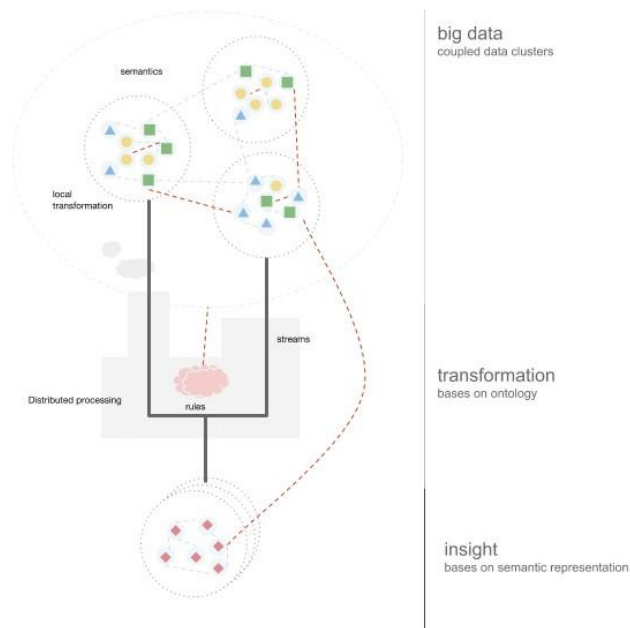
When applied well you could easily combine endless amounts of datasources and query information like: “select all persons which are born between 1900 and 1910 and died on the age of 20 and lived in Italy or Spain”. This is a very flexible solution to obtain any information. It transforms the web into a unified database.

Using semantic technology will save a huge investment in new infrastructures, reduce maintenance costs and no proprietary lock-in.

3 Complement traditional big data approaches with semantic technology

Complimenting the traditional big data approach could also be an option. Below is an image which depicts this approach.

There are several benefits for combining semantic technology and traditional big data approaches.



The quality of data will be much better when having explicit semantics. In addition, both a uniform interface and an uniform data structure is provided due to the coupled clusters. This makes your processing pipeline much more transparent.

Another possibility is to use semantic technology as intermediate format. i.e., translating varies data structures to a semantic formal model and do all kinds of reasoning on the newly created semantic information.

Similarly, the generated representations could be stored as semantic data. In this case representation could link back to the original source for further analysis without losing the context.

4 summary

There are varying ways to get insight to Big Data, one way is using semantic technology. Using semantic technology on its own reduces complexity and resources. Complimenting semantic technology with traditional approaches is also beneficial.



About Taxonic

We are Taxonic, an international and independent IT consulting firm. Our focus? Digital transformation and Knowledge Management, which we apply within two areas: Linked Data and Pega (Dynamic Case Management).

As a knowledge partner, we share our expertise, provide tailored advice, and are hands-on in projects from analysis to implementation. Our experts work with the latest technologies. Within the team, we have a broad range of experience; from advising on technology selection and integrating solutions into the technical environment to connecting the system landscape and complete implementation projects.

We offer a full end-to-end service that allows us to guide digital transformations from design to implementation, ongoing development, and maintenance. Additionally, we are actively involved in developing our own software products. We create innovative software solutions based on Linked Data and semantic technology.

Working with interdisciplinary teams enables us to tackle a wide spectrum of complex projects. Taxonic encourages the world to adapt to new technologies, and this is our contribution to society.

Veldzigt 2
3454 PW De Meern
The Netherlands

T +31 (0)88 829 66 00
E info@taxonic.com
W www.taxonic.com

bank NL45RABO0161959660
VAT NL851339803B01
CoC 54529190