TAXONIC

Big Data processing and Semantic technology a symbiosis

Big Data processing and Semantic technology a symbiosis

A Taxonic whitepaper, by Taxonic B.V., Utrecht, The Netherlands

http://taxonic.com

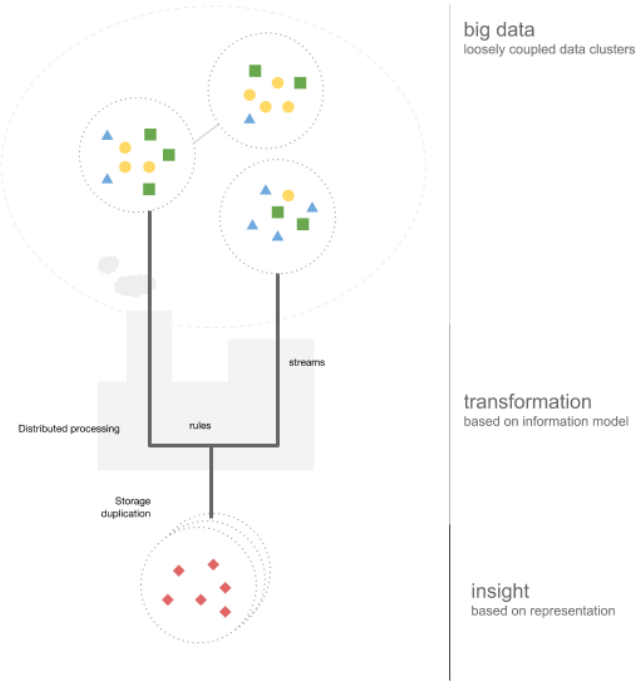
Semantic Technology and Big Data are two abstract ideas. In this article, I present, in high-level, how semantic technology relates to big data and how combining the two yields to better results. For an introduction to semantic technology, Taxonic's whitepaper *Linked Data in the Enterprise* by Jan Voskuil, is a great read!

Big Data

The core idea of big data is to collect data points and make sense out of it. With sufficient data point correlations may be found which helps analysing 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 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.

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 consumptions less of a pain. All the resource challenges you have with traditional approaches are 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.

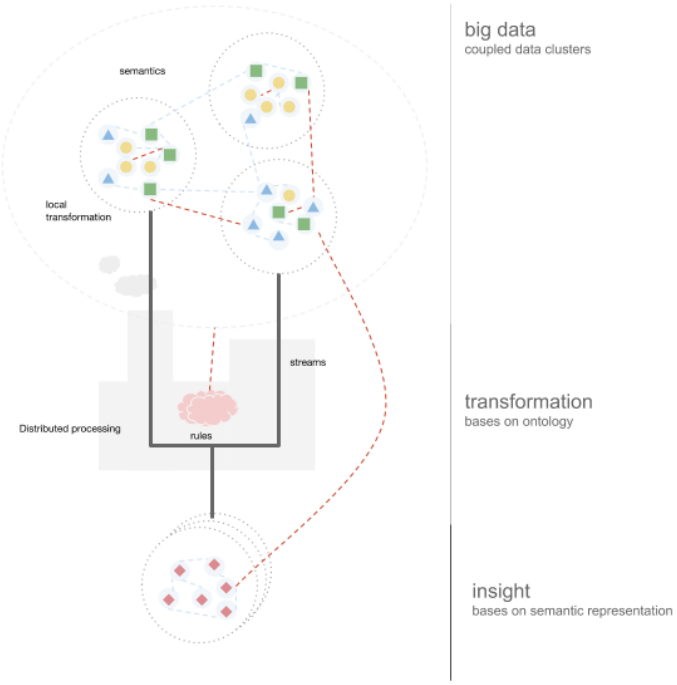
When applied well you could easily combine endless amounts of datasources and query information like: "select all persons which are born between 1900 and 1910 and died on the age of 20 and lived in Italy or Spain". This is a very flexible solution to obtain any information. It transforms the web into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

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

Complement traditional big data approaches with semantic technology

Complement 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 an 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.

summary

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

About Taxonic

Taxonic helps organizations creating more value from their information flow. Linked Data-technologies are an essential ingredient of that. We consult in technology selection, provide expertise in tendering, and assist in the design, delivery, implementation and exploitation of solutions.

Taxonic BV

Janssoniuslaan 80

3528 AJ Utrecht

The Netherlands

T +31 (0) 88 240 42 00

info@taxonic.com

www.taxonic.com

kvk 54529190

rabobank 161959660

btw NL851339803B01