Managing Big Multidimensional Data

Torben Bach Pedersen

Aalborg University
Department of Computer Science
Center for Data-intensive Systems (Daisy)
tbp@cs.aau.dk
http://people.cs.aau.dk/~tbp

Business Intelligence Versus Big Data

Multidimensional data has traditionally primarily been used for Business Intelligence (BI) applications, so it is interesting to check the similarities and differences between BI and Big Data. We now define the two terms.

Interestingly, the term Business Intelligence was coined in 1958, more than 56 years ago. Here, H. P. Luhn defined BI as "the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal" (Luhn, 1958) More recently, in 2013, Gartner Group defined BI to be "an umbrella term that includes the applications, infrastructure and tools, and best practices that enable access to and analysis of information to improve and optimize decisions and performance" (Gartner Group, 2013). Thus, we can conclude that optimizing your business using data is not a new idea.

The term Big Data is of course more recent, although the first uses of the term seems to date back to the 1990's. Wikipedia defined Big Data to be "an all-encompassing term for any collection of data sets so large and complex that it becomes difficult to process them using traditional data processing applications" (Wikipedia, 2014). So, the data should be so "big" that it becomes "difficult" to do it the traditional way.

When defining what is means to be "big", the 3 V's of Volume (very large datasets), Velocity (data arriving very rapidly), and Variety (data of very different format and structure) are well-known. However, up to 6 extra V's are also mentioned, namely Veracity (how much can we trust data ?), Visibility (data must be visible to the Big Data processes), Variability (the meaning of data changes over time/place/context), Viability (can our data be used for anything useful ?), Value (what real value can this data add to our business ?), and Visualization (complex visualization is needed to fully understand and get value).

Now, we can finally compare BI and Big Data. The following issues are *similar* for the two concepts, and thus not novel for Big Data :

— Collecting, integrating, and analyzing data to gain knowledge
— Large data volumes
— Data (often) arrives at a fast pace

Certainly, the first item is what BI was always all about. Traditional data warehouses have also held very large, and increasing, amount of data, and the concept of real-time data streams has