Big Data: Is It Worth It?

This worksheet serves to guide you through deciding whether "big data" technologies are something you actually need. They come at significant costs, and it's not something you should take on just because it's trendy.

What sources of "big data" are generated by your company?

It's best to start by cataloging whether you have "big data" at all – remember the four V's: volume, velocity, veracity, and variety. Not all companies have "big data" to begin with, and can make do with more established relational database technologies for the forseeable future.

Possible sources of "big data" would include logs from large websites, financial transactions, sensor data, claim data, data generated by the activity of your employees and customers, and medical data. Think broadly about your own business here.

What ideas do you have for monetizing that data?

Think creatively! What new products would the analysis of this data make possible? Could you just do a measurably better job at what your business already does by extracting meaning from this data? If so, how much better?

What are those ideas worth?

Think back-of-the-envelope. What's the potential dollar value of implementing those ideas on an annual basis?

What would the cluster to enable those ideas cost?

You'll need to start with an estimate of just how "big" your data is. Remember a single computer can store several terabytes of data; you only really need a cluster if you are generating much more than that over time, or if you want to analyze that data in complex ways that a single computer can't handle.

About how many computers do you think it would take to manage the data you have in mind? Now, multiply that by three, because you need to have redundancy in place as well as additional hardware layers to manage your cluster.

Multiply that by the total cost of ownership (TCO) of a computer in your company. This can be surprisingly high, on the order or $20,000 per year per host once you consider the labor and network infrastructure required to support each computer.

So, is it worth it?