Critical Thinking Option 1

John Wensink

MIS480 Capstone

Colorado State University-Global Campus

Chris den Hejer

November 15, 2020

Critical Thinking Option 1

In 2020 more than ever, it is essential for businesses to operate efficiently and effectively, making use of the modern big data environment to gather high-quality business intelligence for which to guide decisions. The business intelligence framework as it pertains to the manufacturing of goods can be optimized by paying careful attention to the way data collected, stored, analyzed at each stage of the manufacturing process. The best way to get employees motivated to handle data carefully is to familiarize them with the impact their role has on the organizational big picture. Tools exist to help streamline the process of visualizing big data concepts that is simple and straightforward for typical employees to consume. The COVID-19 pandemic has fundamentally changed the way that businesses operate. A focus on digital connectivity means firms must rely on web connectivity like never before. Our organization must be ready to adapt to this challenge and gain the confidence of our customers by providing our products in a highly reliable manner, being prepared for sudden surges in demand as more parts of the world are seeing a spike in new coronavirus cases. Preparing for operational efficiency in a catastrophe does not happen for free. Our organization must seek out efficiencies as they pertain to data mining and business intelligence.

**Developing Business Intelligence**

It is imperative that your organization develop both tactical and strategic roadmaps

toward stated goals, so that success may be measured. As it is now, your organization is not making use of the robust data available to it. Perhaps it is that it is such a large dataset that it seems unmanageable to any one person. Thankfully, tools have been developed which make use of advances in computational resources both in their individual power, and their ability to connect with more powerful systems. These tools need not cost your organization for software or licensing, as there are several open-source options available that are intuitive for a budding data department. Hire an employee who knows his or her way around a relational database. Skills in computational languages such as Python, R, SAS, and other systems will allow your data employee to begin to set up a database for your organization. It is important not to cheap out on disk drives for this critical infrastructure. We will be setting up a business information system that is constantly in operation, reading, and writing data to drives 24 hours a day. Network Attached Storage (NAS) disk drives are a touch more expensive but are rated to be set up in a RAID5 configuration, and spun constantly (Rahman & Shavier, 2018). It should be noted that there have been disk drives manufactured recently which make use of Shingled Magnetic Recording technology which are marked for NAS systems, but in actuality will have read/write issues without being given an amount of downtime in which to defragment striped data (Patana-Anake, Martin, Sandler, Wu, & Gunawi, 2016). Now that you have a NAS system setup to store your data, it is time to begin collecting the information which we will eventually use to make business decisions. Garbage in garbage out is common parlance to describe the effect that poor quality of data inputs will have on our modeling efforts. As we are starting a new system from scratch, let’s be careful to sanitize and preprocess what we put into our tables before we put it in, as it will save you a headache down the road when you are creating models, migrating databases, and validating results. Each and every employee needs to be motivated to begin taking high-quality data in their individual roles. Rules need to be established for atypical and null data points so that they are obtained using a commonly used system. This means taking time to ensure compatibility with all types of devices from Windows, Mac, Linux operating systems, and other types of data sensors and systems. Open-source software is a wonderful way for your organization to start its data department, however, there are serious limitations as it applies to scaling up. Product support, while robust, is likely self-service only which is fine for a single employee, but as your department grows you may wish to consider proprietary data software such as SAS or Tableau. When you trust that your data is clean, secure, and redundant you can start an exploratory analysis of what kinds of data make up your tables. What kinds of relationships are you interested in? Do you notice any patterns? If so, it’s time to make and test hypotheses using all sorts of analysis techniques. There are different techniques for different applications both in describing and predicting various business phenomena. Machine learning algorithms come in a multitude of varieties depending on the types of data involved, modeling outcomes, population/sample requirements, and difficulty of use. A good place to start would be some linear/logistic regression techniques, Bayesian classifiers, decision trees, and even simple artificial neural networks are within your reach at this stage. The only limit is your own creativity, and the regional laws which dictate which types of data are fair game. Your datasets should be vast, meaningful, accurate, and recent. It is our recommendation that you begin work on a data warehouse quickly as your storage space begins to come at a premium. Data warehouses are great places to store data that will not have read/write operations taking place while in storage. They are cheaper to use per unit of storage than your relational database will be once it begins to fill up with data. Different data warehouse solutions are available in different configurations of physical and cloud-based storage architectures. Be careful, and remember, ‘the cloud’ is just someone else’s computer. We expect your curious data scientist(s) to start correlating previously unknown phenomena, and because you were so careful to only input quality data, you can be confident that if your models are sound, your organization will have an advantage over your competitors who are not taking this approach. The field of data analytics is complex and it will not be perfected overnight. Practice makes perfect, you will likely see dramatic results in the early phases of data analysis, cutting the most egregious inefficiencies as they become apparent. Later victories will come at a higher cost, but we believe that the benefits of investing in a robust data analysis program will far outweigh the costs.

References

Patana-Anake, T., Martin, V., Sandler, N., Wu, C., & Gunawi, H. S. (2016). Manylogs:

Improved CMR/SMR disk bandwidth and faster durability with scattered logs. *2016 32nd*

*Symposium on Mass Storage Systems and Technologies (MSST)*. Retrieved November 15, 2020, from <https://www.semanticscholar.org/paper/Manylogs%3A-Improved-CMR%2FSMR-disk-bandwidth-and-with-Patana-anake-Martin/66cdd2d92eef3290cbc06019220e31a63b76660d>

Rahman, P. A., & Shavier, G. D. (2018). Reliability model of disk arrays RAID-5 with data

striping. *IOP Conference Series: Materials Science and Engineering,* *327*, 022087. Retrieved from <https://iopscience.iop.org/article/10.1088/1757-899X/327/2/022087/pdf>