A Value Based Big Data Maturity Model

Badie Farah
Eastern Michigan University

Data has the power to change every business and organization. The issue is what data, and what analysis is applied to that data. Data may be looked upon as a store of information useful for answering questions, providing prospective on problems, and shedding light on potential solutions. Because of this potential, organizations and businesses are enticed to acquire, maintain, and process volumes of structured and unstructured data. This paper advances a big data maturity model which is used to describe the standing of an organization with respect to its big data effort, and provide a road map for advancing such an effort. This model is differentiated from others by its orientation toward the value its big data strategy provides to the organization.

INTRODUCTION

Data has the power to change every business and organization. The issue is what data, and what analysis is applied to that data. A tremendous amount of all kind of data (text, numeric, picture, audio, and video) is generated, stored, and processed daily by organizations, businesses, and individuals. This data may be looked upon as a store of information useful for answering questions, providing prospective on problems, and shedding light on potential solutions. Because of this potential, organizations and businesses are enticed to acquire, maintain, and process volumes of structured and unstructured data. In essence, the field known as Big Data (El-Darwiche, et. al., 2014)

Today's organizations look at big data as a strategic perspective for establishing and enhancing their decision making processes. These newly aspired to enhance their decision making processes are driven by a comprehensive and long term data acquisition, maintenance, and analysis. The current Information Technology (IT) allow the large volume of data, acquired from a variety of sources, to be processed at a high speed. A major question any organization ought to pose before adopting this technology is that: are we paying for valuable data or data plus noise; or only noise? Mapping valuable data with analytical tools provide value for the organization. However, mapping noise with analytical tools might provide confusion which, in turn, might lead to bad decisions. That answer, to the question of the value of the data, will determine if the organization will benefit from this technology or not. And since big data should be a multidisciplinary effort in the organization, answering such a question will involve people from the various functions of the organization. In other words, the IT people should only contribute to providing the answer, rather than providing the answer.

To determine if big data is valuable, the organization should consider the data scope, data quality, and data accessibility.

Data scope: More data is not necessarily the answer. The right data is! Determining the scope of data should be multidisciplinary – multifunctional. This effort is crucial to the success of the big data project.

Data accessibility: What good data is if you cannot access it? Since data might be originating from a number of sources (internal and external), the issue of access to the data must be resolved. It is of great importance that trust of the sources of the data, and the ability of the recipient to verify the data are established. Data sharing and protection of data providers are essential for obtaining valuable data.

Data quality: To ensure data quality, standardization to link data sets is essential. Data integrity is another facet of data quality. Here we are asking how much "bad data" can pollute "good data; how can the organization determine "false positive" and how much such determination cost. On the average 80% of time and effort is spent on cleaning and organizing data while 20% and effort is spent on collecting data.

To ensure confidence in the data and the results of the analysis, the organization may insist on a quantitative measure of the value of bid data, rather than just depend on the force of the argument that the big data proposal will return value to the organization.

Big data is a multi-faceted technology, it is made up of old and new technologies, the objective of which to assist organizations in consolidating and interpreting data for the purpose of achieving competitive edge. This competitive edge may be attained through insight generated by big data technologies by augmenting internal and external data with powerful analytical tools. These technologies provide a platform for managing the load and storage aspects of the organization data. In addition, these technologies make it possible to acquire, manage, and analyze tremendous volumes of data in a timely fashion and provide insight to the organization as to what the analyses has discovered (Halper, et. al., 2013).

BIG DATA MATURITY MODELS

A maturity model for big data is a way of classification to describe the status of an organization with respect to its effort as it relates to data collection, maintenance, and analysis. Maturity models for big data may differ in name, number and names of their various levels, and the content of each level. In the following sections we will summarize some of these models.

One maturity model, unlike others, concentrates on the organization's architecture when describing the status of its big data structure. This model describes how the data is organized to integrate systems for the different applications (Mazzaferro, et. al., 2013).

Another maturity model concentrate on data quality management. Data quality is crucial for the competitive advantages of the business. The better the data quality, the better the information that is generated from analysis of the data, the better the decision that management can make. The competition in business is transformed, in this era of big data, to competition for getting quality data and employing the best powerful analysis tools to transform the data into information for competitive advantage. Maintaining high quality data could be fairly expensive for the enterprise. However, when contrasted with low quality data, it is obvious that it is worth the cost. Low quality data brings a number of negative ramifications to the business including, but not limited to, dissatisfied customers, higher cost for maintaining the data, degraded performance, and ineffective decisions. In addition, low quality data has great effect on the organization's employees moral and trust of their management. A big data quality model focuses on the quality of the data structure and the management of such a structure. This leads to a model for managing a quality metadata model (Ryu, et. al., 2006).

Today's organizations consider big data the latest advancement in Information Technology (IT) that might have a long-term effect to direct and enhance their data-driven decision-making. Big data is characterized by the extraordinary large data volumes, which originate from a variety of sources, with real-time data collection, storage, and analysis (El-Darwiche, et. al., 2014).

Big data represents the most recent advances in data collection, analysis, and usage toward the objectives of the organization. The magnitude of acquired data is massive and the trend for this acquisition is exponential. This necessitates that the organization is able to utilize the data for the benefit of the organization in a timely fashion. A maturity model that is built and administered around the benefit and cost, for the organization, of being at each level is of great help in determining what data to collect, what tools to use in analyzing this data, and what decisions this analysis will facilitate. In other words, a big data program should be based on economic feasibility of the applications affected and their dependent on such a program (Farah, 2015).

In this paper we will develop a benefit/cost/risk maturity model that encompass benefits, cost of data, and risk of not acquiring such data and being at a given level of the model. In addition, we will provide a mathematical model that encompass the benefit, cost, and risk values of the organization applications being at a certain level of the proposed maturity mode. These values will also be presented in a matrix form that provide a snapshot presentation to the decision maker.

A VALUE BASED MATURITY MODEL

This maturity model which we call "A Value Based Maturity Model", like all other maturity models, consists of several levels each of which denoting the commitment of the organization to its data. The proposed model has five levels – Initial, Defined, Managed, Optimized, and Strategic -- which are detailed in the following sections. At this point of the development of the model we found that five levels will serve the purpose of identifying the important activities the big data might impart on the wellbeing of the organization. The names given to these levels imply the commitment of the organization to big data at each level. Other names may be contemplated, as well as, more or less levels depending on the organization needs or desired classifications.

However, unlike other maturity models discussed in the literature, our proposed model is oriented toward the benefits, costs, and risks applications might add to the organization. Because organizations commitment to big data should be dependent on the added value that applications bring to the organization, the proposed maturity assigns a level to any application based on these values. Therefore, a given level may be suitable for some applications depending on their needs of data, and their move to a higher level in the maturity model is based on their requirements of more sophisticated data technologies – such as big data.

Level 1 – Initial (Pre-Contemplation)

This level is characterized by somewhat lack of awareness or denial of the value that a big data program might contribute to the wellbeing of the organization. It also characterized by an ad hoc approach- on the initiative of some people in the organization- to data collection and analysis. Personal initiatives, and use of Data Exploration/Visualization tools for ad hoc analysis of accumulated data are examples of such efforts at this level of the maturity model.

Data Base Management System (DBMS) is probably the only available tool for the accumulation and crude analysis of data at this level. The activities undertaken at this level are oriented toward monitoring the business activities and initial assessment of performance of these activities. In other words, Level 1 utilizes data that is generated by the operation of the organization and maintained by the organization DBMS. This technology is very much available to all organization with even a small commitment to Information technology (IT).

Consciousness raising of the value of big data is essential to get the organization to experiment with this technology for the benefit of the organization. How successful is this endeavor is highly dependent on the benefits the ad hoc projects, that have already been taken, returned to the organization.

Level 2 – Defined (Contemplation)

This level is characterized by awareness of the need to change, by thinking about big data and its benefit to the organization. Contemplation is characterized by adding unstructured data, real time feeds

from external organizations, and feeds from the social media to the data generated by business operations -- which is utilized in the previous level (Initial). This additional data is augmented with statistical tools and data mining software to assist in predicting possibilities for the organization. These possibilities are then feedback into the strategic and operational decision of the enterprise to generate more efficiencies and new products and markets.

Organizations at this level need more conscientious raising. This may be achieved by reading, studying, and engaging their specialist in seminars and conferences about big data. Being exposed to big data in this fashion promotes awareness of potential problems, cost, risks, and benefit of big data to the organization.

Level 3 – Managed (Preparation)

In this level the organization is preparing to change its current ways and adopt big data technologies for the management of its data throughout its applications. This change is about to take place very soon since resources have been committed to deliver this change. The organization, at this stage, is forward looking to the future to harness the power of big data. Level 3 is characterized by deploying Metadata Systems, and integrating and managing data across all applications. Data integration across application can deliver great benefits to the organization by shedding light on the possibilities that might arise from integrating applications. The commitment of the organization to big data at this level is significant in monetary, as well as, personnel and organizational resources.

Organizations need to avoid making the mistake of expecting miracles, in monetary returns or improvements in decision making, in a short period of time. Therefore, they must set realistic goals to avoid disappointment and potential failure of this undertaking.

Level 4 – Optimized (Commitment)

This level is characterized by the need for a higher level of data management. Enterprise Data Management Architecture becomes essential at this level because it provides an organization wide view of the data which permeate the applications. A top level view of the data, which can be articulated in terms of the organization goals and objectives, and further details to provide views appropriate at the application level, can all be accommodated by the enterprise Data Management Architecture.

The organization must also take on the tasks of optimizing data collection, scheduling, and workload. Furthermore, data security and data governance are at the center of dealing with the organization data at this level.

Level 5 – Strategic (Future)

This level is characterized by the foresight that big data is one of the great asset of the organization and it is to be exploited for strategic advantage. Accumulating data and subjecting it to the best analytical tools available for extracting trends and new venues to pursue new products, markets, and ventures becomes a strategic objective of the organization.

The organization, at this level, is looking to leverage its big data for new revenue opportunities. The major concern should be centered on the value that the big data could add to the net profit of the organization. In other words, the cost of deploying big data, and the risk of not deploying big data, should be a major consideration in the adoption decision of these technologies for the proposed applications.

The organization should shadow its customers and research their needs and desires. In other words, the organization should move its data collection and analysis from just a product/service orientation to a platform where it looks for a more futuristic orientation.

A Benefit/Cost/Risk Model for Determining the Appropriate Level

This is a business based model which may be used to determine the appropriate level at which an organization should be placed in the proposed maturity model. The model assesses the costs, benefits, and risks associated with applications being placed at a certain level of the maturity model. The model we advocate is:

$$B(i) > C(i) + R(i) \quad (1)$$

Where

B (i) denotes the benefit to the organization of being at level i in monetary unit say $;
C (i) denotes the cost to the organization of being at level i in monetary unit such as $;
R (i) denotes the risk to the organization of NOT being at level i in monetary unit say $.

C (i) is the cost that the organization will incur in being in level i of the maturity model. That represents all the cost of technology –including hardware, software, and human expertise—necessary for moving the organization from level i-1 to level i.

R (i) may be represented by the opportunity loss for the organization that result from staying at level i-1 and not moving to level i. Other representation and assessment such as penalty that might be charged to the organization because not being able to confirm to regulations or standards. This might be the case when being at level i-1 does not provide the organization the necessary data or analysis that are essential for compliance.

The model in equation (1) may be expressed in terms of the B, C, and R components of the individual application in use by the organization. Such an expression is depicted in the following relation:

$$B(i, j) > C(i, j) + R(i, j) \quad \text{for all } i,j \quad (2)$$

Where

B (i, j) denotes the benefit to the organization of having application j in level i in monetary units say $;
C (i, j) denotes the cost to the organization of having application j in level i in $;
R (i, j) denotes the risk to the organization of NOT having application j at level i in $.

Unlike other big data maturity models, this model is built on the economy of having a particular organization at a certain level depending on the net benefit, of being at such a level, to the organization. Furthermore, associating the net benefit with each application provide more understanding as to why, or why not, the organization needs to move into a particular level of the maturity model.

The model in relation (2) can be summed over all i, j to achieve an equivalency to the model in relation (1). That is:

$$\sum B(i, j) > \sum C(i, j) + \sum R(i, j) \quad \text{for all } i, j \quad (3)$$

Several ratios from the above variables and sums may be calculated to shed more light on the effect of having a particular application at a particular level of the maturity model. Some of these ratios are given in the following relations with respect to level i.

$$B(i, j) / \sum B(i, j) \quad \text{for all } j \quad (4)$$

Relation (4) describes the benefit contribution of placing application j at level i when measured against the benefits of all applications at level i.

$$C(i, j) / \sum C(i, j) \quad \text{for all } j \quad (5)$$

Relation (5) describes the cost burden of placing application j at level i when measured against the cost burden of all applications at level i.

$$R(i, j) / \sum R(i, j) \quad \text{for all } j \quad (6)$$

Relation (6) describes the risk associated with placing application j at level i when measured against the total risk of all applications at level i.

Another set of ratios may be computed from relations (1), (2), and (3). These ratios facilitate the decisions with respect to a big data program where a multiple of applications feed from such a program. Some of these ratios are given in the following relations with respect to all applications that benefit from a big data program.

$$\sum B(n, m) / \sum B(i, j) \quad \text{for all n, m of applications that benefit from the big data program; and for all i, j applications of the organization} \tag{7}$$

Relation (7) describes the benefit contribution of all application at all levels that benefit from a certain big data program when measured against the benefits of all applications at all levels.

$$\sum C(n, m) / \sum B(i, j) \quad \text{for all n, m of applications that benefit from the big data program; and for all i, j applications of the organization} \tag{8}$$

Relation (8) describes the total cost burden of all application at all levels that use a certain big data program when measured against the total cost burden of all applications at all levels.

$$\sum R(n, m) / \sum B(i, j) \quad \text{for all n, m of applications that benefit from the big data program; and for all i, j applications of the organization} \tag{9}$$

Relation (9) describes the total risk associated with all application at all levels that use a certain big data program when measured against the total risk of all applications at all levels.

The model in equation (2) may be represented in a matrix form where the rows refers to the levels of the maturity model and the columns refer to the application. The body of the matrix holds the corresponding entities of B, C, and R (See Table 1).

TABLE 1
A MATRIX REPRESENTATION OF THE COST/BENEFIT/RISK MODEL FOR DETERMINING THE APPROPRIATE LEVEL

Application j / Level i	1	2	3				j	∑i,j
1							B (1,j), C (1,j), R (1,j)	∑B (1,j) for j, ∑C (1,j) for j, ∑R (1,j) for j
2		B (2,2), C (2,2), R (2,2)						
3								
4								
5								

This table provides an easy to read and compare format of the values of the Cost/Benefit/Risk model for the various levels and applications. This will help the decision maker decide on whether, or not, to dedicate the necessary resources for big data technologies for a given application. In addition, and at a glance, the decision maker sees which application is contributing to the benefit that accrue from the big data endeavor at the organization.

The last column of the table includes sums across the row for all B, C, and R over all the columns of a given row. The sum provide a snapshot of the values of B, C, and R for all applications at a given row. The decision maker can compare these values across all levels to determine the contribution, cost, and risk associated with the various levels and their corresponding applications.

Depending on the level of the application with respect to the maturity model, the organization might have computed estimates of B, C, and R for the application. If not, the organization might have procedures, in place, to calculate such estimates. In other words, estimating values for B, C, and R should not be a great burden for the organization particularly if it had used a maturity model in the past.

SUMMARY

Maturity models for big data serve as a tool to focus the organization's activities on the best possible outcome for its efforts in acquiring, maintaining, and manipulating data. The current level of the organization's big data effort tells of the possibilities of future benefit to the organization if it pursues the next level of the maturity model. So in essence a maturity model is a road map that an organization can adopt to guide its future efforts, if any, in big data. This type of maturity model serves as a classification tool to assign a level to the current status of the organization big data program; and show what future levels require in terms of cost and risk, and what they would return in benefit to the organization.

In this paper we propose a value based model, which in addition to serving as a classification tool of the organization's big data effort, takes into consideration measures of cost, benefits, and risk associated with each level of the maturity model. This model provides the organization a valuable tool that can be used to pursue big data technologies based on the value they add to the wellbeing of the organization. A number of mathematical relations assist in determining the appropriate level for applications by calculating the costs, benefits, and risks associated with the various applications that are implemented by the organizations and considered as candidates for big data technologies. Some ratios of benefits, costs, and risks are also calculated. These ratios provide measures of the incremental values – positive or negative -- big data applications contribute to the organization. Furthermore, a matrix representation of this model is provided. This matrix representation provides a snapshot of the standing of the organization in terms of the levels and applications. The cells of the matrix contain the cost, benefit, and risk associated with each application in its current level. As applications move from one level to another, the matrix graphically demonstrate the past and present standing of the applications with respect to big data maturity model.

REFERENCES

El-Darwiche, B., Koch, V., Meer, D., Shehadi, R.T., & Tohme, W. (2014). Big Data Maturity: An Action Plan for Policymakers and Executives. *The Global Information Technology Report,* 43 – 51.

Farah, B. N. (2015). Big Data – What Data and Why?. *Journal of Management Policy and Practice*, Vol. 17, Issue 1, pages 11 – 17, 2016.

Halper, F., & Krishnan, K. (2013). TDWI BIG DATA MATURITY MODEL GUIDE; Interpreting Your Assessment Score. *TDWI Research*, 1–16.

Hogan, M., Liu, F., Sokol, A., & Tong, J. (2011). NIST Cloud Computing Standard Roadmap. *NIST Special Publication.500-291.* 2*,* 1-102.

Hüner, K.M., Ofner, M., & Otto, B. (2009, March). Towards a maturity model for corporate data quality management. *Proceedings of the 2009 ACM symposium on Applied Computing.* 231-238. ACM.

Mazzaferro, G. (2011, November). Towards A General Reference Architecture for BIG DATA Creating a Formal Specification. *AlloyCloud*. 1-41.
Pednault, E. (2011). Big Data Platforms, Tools, and Research at IBM. *IBM Research*. 1-35.
Ryu, K.S., Park, J.S., & Park, J.H. (2006). A Data Quality Management Maturity Model. *ETRI Journal. 28(2),* 191-204.