https://doi.org/10.48009/1_iis_2015_200-209

Issues in Information Systems
Volume 16, Issue I, pp. 200-209, 2015

BIG DATA AND BUSINESS ANALYTICS IN A BLENDED COMPUTING-BUSINESS DEPARTMENT

Michael G. Hilgers, Ph.D., Missouri University of Science & Technology, hilgers@mst.edu
Sarah M. Stanley, Ph.D., Missouri University of Science & Technology, stanleysm@mst.edu
Cassandra C. Elrod, Ph.D.* Missouri University of Science & Technology, cassa@mst.edu
Barry B. Flachsbart, Ph.D., Missouri University of Science & Technology, barryf@mst.edu

ABSTRACT

The need for trained graduates in various aspects of "big data" is noted and three roles are identified: Data Scientist, Data Analyst, and Data Explorer. The activities that each of these specialists is likely to undertake is specified and the quantitative, computer, and business skills that each would require is noted. Traditional business and information science/technology (BIT) degrees are examined, including especially degrees from hybrid combined business and technology programs. After examining the details of quantitative, computer, and business skills required, it appears that BIT graduates require additional skills. Three hybrid courses that provide additional strength in the skills needed are proposed: Introduction to Data Science and Management, Business Analytics and Data Science, and Data Science Programming in Python. By adding these courses, it is believed that BIT degree graduates could operate as data analysts or data explorers and contribute to filling the needs for big data expertise in industry.

Keywords: Big Data, Business Analytics, Curriculum Development, Python

INTRODUCTION

It is no secret that customers have gone digital. They are connected on so many levels through the use of smartphones, tablets and gaming consoles. This trend is producing an inordinate amount of data that can be used to better communicate with and attract more customers [16, 5, 8]. Utilizing the data in a profitable manner is the new challenge of this computing age. Referred to in umbrella terms as the "big data" problem, capturing and analyzing diverse data sets is creating demand for people with a different skill set than has been the norm in the computer/information science fields [19]. The purpose of this paper is to examine the coalescing functional roles within the big data environments, clarify their defining skills, then decompose such skills in terms of common academic coursework. It will be seen that traditional business and information science/technology degrees (for brevity denoted as BIT degrees) need to be enriched with certain subjects to remain competitive in placing its graduates. This paper then proposes that by adding a few hybrid, cross-disciplinary courses to such programs, they can sufficiently enhanced to produce graduates who are competitive in the big data world.

BACKGROUND

Recently a sub-group of the Association for Information Systems called for universities to direct more resources at preparing students for careers in big data, and the revamping of curriculum to keep pace with industry [19]. Research suggests that in order to better prepare students, universities need to become more cross-functional and use a cross-disciplinary approach to prepare students for the jobs they will be filling upon graduation [2, 12]. There is, however, still confusion over the exact definition of big data and where it should be 'housed' at a university [18]. Weinberg et al. argue that big data can be condensed down to two groups: technology focused and managerial problem solvers. The technology focused consists of traditional academic areas such as computer science and computer engineers while the managerial problem solvers are more likely to reside in the areas of marketing and management [18]. Yet other researchers believe that there are four characteristics of big data (1) it comes from multiple sources (2) it is unlikely to be owned by any one source 3) it isn't managed by traditional database tools (4) it is 'gigantic' [1].

Similar to the defining "big data" is the need to solidify the concept of "business analytics," which is emerging as the size of data sets grows. Business analytics involves several phases and requires a number of different types of people with complementing skill sets to capture, report, predict, act, and refine the data; turn it into information; and then solve business problems with the information obtained [6, 9]. Others define the major steps of business analysis workflow as acquisition, extraction, integration, analyzing, and interpretation [17].

To highlight the significance of the "big data" problem, a recent study showed that within a few years just the United States alone will encounter a shortfall of 140,000 to 190,000 people with "deep analytical skills." And of even greater significance to the BIT world is a need for 1.5 million "data-savvy" managers who can analyze big data to make effective decisions [4]. With the shift to big data, for example, researchers suggest that all marketers and PhD students have some base understanding of computer science skills and a base understanding of data [18]. They argue that marketers should be learning about big data, not as an outsider working alongside information science or computer science individuals, but instead by taking ownership of this trend which is likely to redefine marketing for many organizations [18]. It is important to note that with this trend, which is helping to define customer experience, there are also many concerns, some of which include adequate training of students to use such data and the ethics involved [9, 14].

How can business schools engage themselves in the opportunities being created? Some schools have achieved success by working with industry. For instance, Wharton has worked with IBM to create a 2-week course targeted to Chief Marketing Officers on the topic of marketing analytics [11]. By becoming more focused on analytics, universities can not only prepare their students better but also boost recruitment. In fact, Brandman University has lowered their cost per acquisition while significantly improving student success by finding the right students. This is evidenced by higher than average graduation rates and an exceptionally low loan default rate for the students they have acquired by leveraging marketing analytics [10].

Functional Roles

Within this environment, industrial practitioners claim that three roles have developed: data scientist, data analyst, and data explorer [7]. Their distinctive roles and responsibilities are given in Table 1.

Table 1. Definition of the Data Scientist, Analyst, and Explorer Functional Roles

Functional Role	Primary Activities
Data Scientist	Has a broad knowledge of business needs and the options available in descriptive, predictive, and prescriptive analytics. Searches, identifies, collects, and transforms data suitable for strategic action
Data Analyst	Has a specific knowledge of the needs of a particular business domain and understands the information required for the mathematical models and statistical analysis to answer specific questions
Data Explorer	Has expert knowledge of particular business functional areas and can formulate questions and find information capable of being compiled in analytical models.

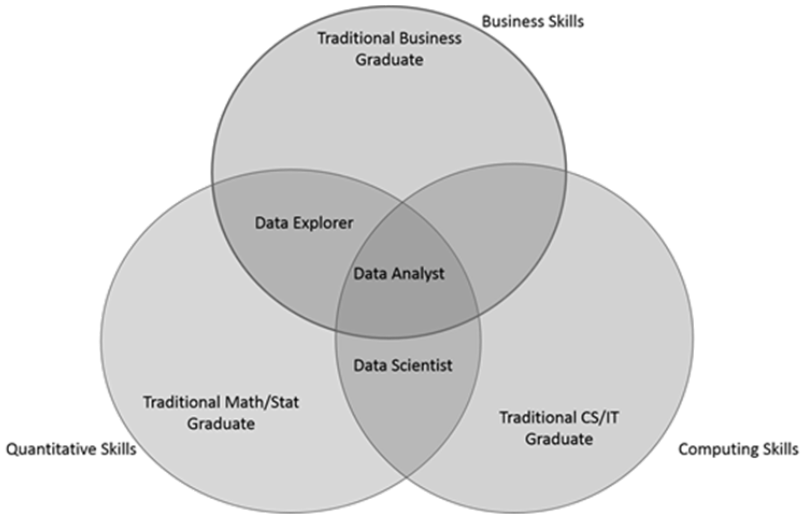

Figure 1. Functional Roles with Distinguishing Skills

Defining Skills

For an educational program to produce individuals capable of entering into these roles, the fundamental underlying skills employed in the execution of their duties must be clearly understood. Only then can curriculums be appropriated modified to maintain pace with the significant changes in the marketplace. Figure 1 offers an attempt to view the new functional roles from the perspective of traditional areas within most university degree programs; namely business, mathematics, and computer science/information technology.

To be a data scientist, one must possess strong skills in computing and mathematics. It is on their shoulders that the most fundamental analysis problems will fail. Meanwhile, the data analyst works in an overlapping world which requires intelligent interaction with computational modeling experts as well as corporate executives. Hence, the need for skills from all of the fundamental areas. The explorers work with the business leaders in the process of imagining the kinds of questions that might be answered to boost the bottom-line. To keep their imagination within bounds, the data explorer must have knowledge of what is possible and of what other analysts have done.

Table 2. Comparison of Quantitative Skills

Quantitative Skills				
Knowledge Areas	**Comparative Need**			**BIT Degree**
Foundational Topics	**Scientist**	**Analyst**	**Explorer**	**Available Content**
Mathematical Modeling	High	High	High	None
Statistical analysis	High	High	Medium	Intro
Model Testing and Evaluation	High	High	Medium	None
Numerical algorithms and methods	High	Medium	Low	None
Visualization	High	High	High	Advanced

Tables 2, 3, and 4 offer perspective on three inter-related concepts. First, an effort is made, by highlighting several foundational topics, to be more specific about the types of quantitative, computer, and business skills needed to operate in the new business analysis functional roles. Based upon experience with teaching and researching in business analytics, these topics provide a rough first pass at such a clarification. To continue with the distinction of the three functional roles in business analytics, the significance is ranked of these foundational topics in the duties of

data scientists, analysts, and explorers. To create the bridge to the topic of BIT curricula, a measure is given of the type of content available (if even obtainable) from these foundational topics to a student in a typical BIT degree.

Table 3. Comparison of Computer Skills

Computer Skills				
Knowledge Areas	**Comparative Need**			**Bus & Info Tech Degree**
Foundational Topics	**Scientist**	**Analyst**	**Explorer**	**Available Content**
Architecture and Networking	High	Medium	Low	Basic
Information Storage and retrieval	High	High	Low	Advanced
Programming Languages and methods	High	Medium	Low	Basic
Big data tools (Hadoop, etc.)	High	High	Low	Basic
Computational Intelligence	High	Medium	Low	None
Data Mining	High	Medium	Low	Introductory

Table 4. Comparison of Business Skills

Business Skills				
Knowledge Areas	**Comparative Need**			**Bus & Info Tech Degree**
Foundational Topics	**Scientist**	**Analyst**	**Explorer**	**Available Content**
Marketing	Low	High	High	Advanced
Strategy	Medium	High	High	Advanced
Finance	Low	Medium	High	Advanced
Customer Research	Medium	High	High	Advanced
Product Development	Low	Medium	High	Advanced
Business Organization	Medium	Medium	High	Advanced

In order to develop the skills in the foundational topics identified in the foregoing tables, universities have the simplest option of finding and utilizing existing coursework. When packaged together appropriately, it is hoped that a passing student will have acquired the skills appropriate for his or her desired functional role in big data problems. As these tables suggest, BIT degree students encounter various difficulties. Perhaps the greatest encountered are within the quantitative skill set. Most BIT degree programs require a single semester each of "business" calculus and statistics whereas most engineering and science majors take three semesters of standard calculus. To complicate matters, "business" calculus and statistics does not provide the prerequisite required for many significant courses in quantitative and computer skillset coursework.

A fortunate trend, however, is that many business and management information systems degrees are requiring courses in information technology [15]. Similarly business courses are being introduced into other more technical degrees. Hence, when referring to BIT majors, it is supposed that business majors take information technology courses, and information technology students take business courses. In fact, some school have gone as far to create a cross-disciplinary programs deemed as information science and technology (IST) or Business and Management Systems (BMS). It is from these hybrid degree programs that Table 3 was completed. IST students take traditional IT courses such as programming, database, networking, and computer architecture. Furthermore, courses in enterprise systems provide a cross-over to business while further developing skillsets in system architecture and structured information storage and retrieval. Unfortunately, IST students do fall short of the advanced topics in computing because they do not have the requisite mathematics/statistics backgrounds or multiple courses in structures, algorithms, and computing theory.

Table 4 paints a very positive picture for students in the BIT community. Most business schools offer advanced material on all of the foundational topics listed. The issue at hand is to open the opportunity for business schools to provide education suitable for the three primary functional roles in business analytics. Table 4 was completed from the perspective of IST or BMS students having the elective option to specialize in one or more of the business foundational topics. As mentioned in the introduction, marketing is a particularly strong choice.

Supporting Coursework

The question at hand is how business and information science degree programs might be enhanced to produce graduates who are competitive in the big data world. So far, three functional roles have been identified in the big data analytical community: data scientist, analyst, and explorer. A first pass through understanding the skills that distinguish the various roles was to classify them as quantitative, computer, and business. To step closer to a BIT curriculum, the skill sets were decomposed into knowledge area foundational topics, then a measure of the presence of these topics was given. It was found that BIT degrees have difficulties with most topics in the quantitative skill set and some topics in the computer skillset, provided attention is restricted to hybrid programs blending business and information science and technology. Now, it is time to be even more specific. What does it mean to say BIT students have no available content in modeling? To clarify this, each of these knowledge area foundational topics must be viewed from the perspective of traditional academic subjects.

QUANTITATIVE SKILLS			
Knowledge Area	**Supporting Subjects**	**Traditional Course**	**Issues**
Mathematical Modeling			
	Structure of a model		IP
	Matrix Algebra	Matrix Algebra	IP
	Nonlinear Functions	Calculus +	IP
	Random Structures	Appl. Probability	TET
	Graphs & Networks	Discrete Math	IP
Statistical Analysis			
	Parametric Estimation	Regression Analysis	IP
	Optimization	Regression Analysis	IP
	Descriptive Statistics	Statistics	TET
	Predictive Statistics	Applied Statistics	TET
Testing and Evaluation			
	Hypothesis Testing	Statistics	TET
	Error Analysis	Statistics	TET
	Stability Analysis	Numerical Analysis	TET
Numerical Analysis			
	Linear System Solvers	Numerical Analysis	IP
	Random Numbers	Numerical Analysis	IP
	Operation Count	Algorithms	IP
	Convergence	Numerical Analysis	IP
	Memory Requirements	Algorithms	IP

Figure 2. Decomposing Knowledge Area Functional Roles into Supporting Subjects (Part 1)

COMPUTER SKILLS			
Knowledge Area	**Supporting Subjects**	**Traditional Course**	**Issues**
Networking			
	Physical Elements	Network Perform. & Design	
	Protocols	Network Perform. & Design	
	Enterprise Systems	ERP	
	Software	Network Perform. & Design	
Information Storage and Retrieval			
	Database Management	Database Management	
	Data warehousing	Data Warehousing	
	Preprocessing	Non-traditional	
	Cloud	Cloud Computing	IP
	Hadoop/MapReduce	Varies	
Programming Languages and Methods			
	Object-Oriented (Java…)	Standard	
	Scripting (R, Python…)	Non-traditional	
	Mathematical (Matlab…)	Non-traditional	IP
	Statistical (R, SAS…)	Non-traditional	IP
	OS Level (Unix…)	Operating Systems	
Computational Intelligence			
	Neural Networks	Specialty	IP
	Evolutionary	Specialty	IP
	Genetic	Specialty	IP
Mining Techniques			
	Data	Data Mining	IP
	Text	Non-traditional	IP

Figure 2. Decomposing Knowledge Area Functional Roles into Supporting Subjects (Part 2)

Figure 2 addresses this step as follows. Each knowledge area foundational topic is decomposed into particular supporting subjects that play a major role. This is subjective. Important subjects may have been neglected; but a survey of textbooks, research, and teaching experiences finds these subjects to be prevalent and often troublesome for BIT students. Once identified, this raises the question of what, if any, traditional academic course might address this supporting subject. The tie to the BIT curriculum is continued by considering what, if any, is the issue in including the supporting subject.

The abbreviations used are "IP" for "Insufficient Prerequisite" and "TET" for "Terminal Elementary Treatment". These are similar in notion. TET suggests that an effort is made to include the supporting subject but it is light in content and terminology in that it does not provide a prerequisite to any course that follows. An example is "business" statistics. IP addresses a situation in which there is no content on the supporting subject and the student does not possess the coursework needed to take the subject.

A survey of Figure 2 offers two key observations. For the computing skills table, the matter is not that grim, provided attention is again focused on IST majors. Most of the supporting subjects are addressed at some level within the degree (or can be taken). Those that are missing in computer skills, unfortunately, are nearly impossible to learn at a detailed level due to the lack of foundational coursework.

Figure 2 again highlights that BIT degrees have a shortfall in quantitative skills pretty much across the table. Fortunately, not all of the subjects require a high-level of mastery for those working in business analytics. For example, matrix algebra is important, but not all of the topics taught in a typical undergraduate course are required. Notation, operations, and system inversion are key, whereas general representation theorems are less so. This offers

hope that some topics can be integrated into a BIT curriculum at a level appropriate for the preparation of the students.

HYBRID COURSES AND RESULTS

One solution to resolving these shortfalls, which is being explored, is to create three hybrid courses. The first course, entitled "Introduction to Data Science and Management," (IDSM) would be required of IST majors and provide an introductory level of mastery on several key topics. This course can be followed by "Business Analytics and Data Science" (BADS) as an elective for those interested in pursuing careers in this area. The topical outlines for these classes are given in Figure 3. In examining Figure 5, it is seen that these two courses address, at least at some level, many of the weaknesses in quantitative skills. These topics are taught assuming a lack of background and re-enforced across two semesters. A third course is used to establish a complete foundation in data science programming languages for BIT majors. To clarify the need within the BIT curriculum, an understanding the professional market is required. A 2014 survey of 719 data science professionals by Gregory Piatetsky [13] found that R, Python, SQL, and SAS are used by 91% of respondents. SQL is already taught in a required database management class. SAS is available by elective in an upper-level data mining course. Furthermore R remains the most significant data science language justifying its inclusion in both required and elective courses. This identifies a need for Python, which is growing rapidly in the business analytics community [3]. Hence, another course is added to the BIT offerings called "Data Science Programming in Python" (DSPP) whose topic list is seen in Figure 4. It provides students with support in Computer Skills as well as Quantitative skills, as is seen in Figure 5.

INTRO TO DATA SCIENCE AND MANAGEMENT		
Data		
	Type Classification	
	Data Structures	
File I/O		
	File Storage	
	File Types	
Summarizing Data		
	Descriptive Statistics	
	Related Visualizations	
Structure of Models		
	Response	Linear (Simple Matrix)
	Factors	Polynomial/Exponential
Parametric Estimation		
	Least Square Estimatio	Prediction
	Quality of fit	
Visualizations		
	Point	Bar
	Line	Including Factors
Programming Techniques		
	Strings	Decision
	Functions	Iteration

BUSINESS ANALYTICS & DATA SCIENCE		
Data		
	Type Classification	
	Data Structures	
File I/O		
	File Storage	API usage
	File Types	Cleaning
Functional Programming		
	Scripts	Data conversion
	String Manipulation	Parsing
Summarizing Data		
	Descriptive Statistics	
	Related Visualizations	
Matrix Algebra		
	Notation	Solving Linear Systems
	Operations	Speed & Memory
Parametric Math Models		
	Linear Regression	Parametric Sensitivity
	Error Analysis	Logistic (Nonlinear)
Nonparametric Models		
	Clusters	
	Nearest Neighbor	
Random Structures		
	Random Number Generation	
	Random Graphs	

Figure 3. Introductory and Advanced Business Analytics Courses

DATA SCIENCE PROGRAMMING WITH PYTHON		
Programming in Python		
	Commandline	Conditional/Iterative
	Data types	Basic Functions
File I/O		
	File Format	
	Storage/Retrival	
Basic Data Science		
	Statistics (NumPy)	
	Visualizations (Matplot)	
Data Preprocessing		
	Regular Expressions	Scraping
	Parsing	Transformations
Supervised Machine Learning (scikit-learn)		
	Parametric Models	Methods (LSE…)
	Regression	Nueral Nets
	Model Selection	Text Mining
Unsupervised Learning		
	Clustering	
	Classification	
Big Data		
	Hadoop	
	MapReduce	

Figure 4. A Data Science Programming Course

QUANTITATIVE SKILLS			
Knowledge Area	**Program Weakness**	**Topic Coverage**	**Relative Intensity**
Mathematical Modeling			
	Structure of a model	IDSM, BADS, DSPP	In-Depth
	Matrix Algebra	BADS	Basic
	Nonlinear Functions	IDSM, BADS, DSPP	In-Depth
	Random Structures	BADS	Introduction
	Graphs & Networks	BADS	Introduction
Statistical Analysis			
	Parametric Estimation	IDSM, BADS, DSPP	In-Depth
	Optimization	BADS, DSPP	Basic
	Descriptive Statistics	IDSM, BADS, DSPP	In-Depth
	Predictive Statistics	IDSM, BADS, DSPP	In-Depth
Testing and Evaluation			
	Hypothesis Testing	BADS	Basic
	Error Analysis	BADS	Basic
	Stability Analysis	BADS	Basic
Numerical Analysis			
	Linear System Solvers	BADS, DSPP	Basic
	Random Numbers	BADS, DSPP	Introduction
	Operation Count	BADS	Introduction
	Convergence	BADS	Introduction
	Memory Requirements	BADS	Basic

Figure 5. Summary of Impact of Hybrid courses on Supporting Subjects (Part 1)

COMPUTER SKILLS			
Knowledge Area Topic	**Supporting Subjects**	**Topic Coverage**	**Relative Intensity**
Information Storage and Retrieval			
	Preprocessing	IDSM, BADS, DSPP	In-Depth
	Cloud	BADS, DSPP	Introduction
	Hadoop/mapreduce	BADS, DSPP	Introduction
Programming Languages and Methods			
	Scripting (R, Python…)	IDSM, BADS, DSPP	In-Depth
	Mathematical (Matlab,Maple…)		Not-Covered
	Statistical (R, SAS…)	IDSM, BADS	In-Depth
Computational Intelligence			
	Nureal Networks	DSPP	Introduction
	Evolutionary		In-Depth
	Genetic		In-Depth
Mining Techniques			
	Data	BADS, DSPP	Introduction
	Text	BADS, DSPP	Introduction

Figure 5. Summary of Impact of Hybrid courses on Supporting Subjects (Part 2)

CONCLUSIONS

Coming full circle back to Figure 1, it has been suggested that by adding a few hybrid courses, BIT degrees can be competitive in the big data business analytics world. It should be observed that BIT graduates would have difficulties operating as data scientists because their level of intensity in computing and quantitative skills remains too low. However, careers as data analysts or explorers seem reasonable.

REFERENCES

1. Andonova, Y. (2013). Loyalty 3.0: How big data and gamification are revolutionizing customer and employee engagement. *Journal of Marketing Analytics, 1*(4), 234-236.
2. Angela D'Auria, S. (2006). Bridging the academic/practitioner divide in marketing. *Marketing Intelligence & Planning, 24*(3), 233-244.
3. Asay, M. (2013). Python displacing R as the programming language for data science. Available: http://readwrite.com/2013/11/25/python-displacing-r-as-the-programming-language-for-data-science
4. Brown, B., Chui, M., & Manyika, J. (2011). Are you ready for the era of 'big data'? *McKinsey Quarterly,* (4), 24-35.
5. Buck, C. D. K., et al., (2014). Mobile consumer apps: Big data brother is watching you. *Marketing Review St. Gallen, 31*(1), 26-35.
6. Campbell, J. P., & Oblinger, D. G. (2007). Academic analytics. *EDUCAUSE White Paper*.
7. Feinleib, D., Big data bootcamp: What managers need to know to profit from the big data revolution. Apress.
8. Ferguson, R. B. (2013). How analytics can transform business models. *MIT Sloan Management Review, 54*(4), 1-8.
9. Gopaldas, A. (2014). Marketplace sentiments. *Journal of Consumer Research, 41*(4), 995.
10. Haire, T. (2015). Educating at Brandman University. Responsive: Multi-channel direct advertising. 23(4), 20-25.
11. *IBM.* (2014). IBM and the Wharton School launch first-of-a-kind program to help chief marketing officers use big data and analytics to deepen customer engagement. *Marketing Weekly News,* 134.

12. Maklan, S., Peppard, J., & Klaus, P. (2015). Show me the money. *European Journal of Marketing, 49*(3/4), 561.
13. Nuggets, K. (2015). Four main languages for analytics, data mining, data science. Available: http://www.kdnuggets.com/2014/08/four-main-languages-analytics-data-mining-data-science.html
14. Nunan, D., & Di Domenico, M. (2013). Market research and the ethics of big data. *International Journal of Market Research, 55*(4), 505.
15. Stephens, C. S., & O'Hara, M. T. (2001). The core Information Technology course at AACSB-accredited schools: Consistency or chaos? *Journal of Education for Business, 76*(4), 181.
16. Svilar, M., Chakraborty, A., & Kanioura, A. (2013). Big data analytics in marketing. *OR-MS Today, 40*(5), 22.
17. Tulasi, B. (2013). Significance of big data and analytics in higher education. *International Journal of Computer Applications, 68*(4).
18. Weinberg, B. D., Davis, L., & Berger, P. D. (2013). Perspectives on big data. *Journal of Marketing Analytics. 1*(4), 187-201.
19. Wixom, B., et al., (2014). The current state of business intelligence in academia: The arrival of big data. *Communications of the Association for Information Systems, 34*, 1.