The success of many different kinds of e-business operations depends on data and information, and how they are used to optimize operations, drive sales and marketing, and grow the business. The ability to manage and safeguard data as a strategic asset, transform it into actionable information, and use it as a strategic differentiator is a key contributor to the success of any business operation. What makes this an even more interesting challenge is the speed at which data have been growing in recent years, due to social networking, the Internet, mobile telephony and all kinds of new technologies that create and capture data. The popular press, McKinsey Consulting, IBM Research and many other organizations are now referring to this development with the phrase “big data”. The broad recognition in industry is that e-commerce operations must be designed to take advantage of the data that have become available, as a basis for improving customer service, achieving firm awareness and making markets effective places for economic exchange. As a result, traditional data management, data engineering, and data analytics techniques do not seem to work well with the historically large amounts of data that many e-commerce operations face today. Indeed, these day electronic commerce is all about data!

Despite the critical role that data play in the success or failure of e-commerce firms, there has not been enough research on effective ways to leverage it to create meaningful information for management and strategy in e-commerce. This special issue is intended to highlight the need for more systematic research and investigation in this important area. The research should include many different types of data, including: quantitative and qualitative data; text, audio and video data; stocks of existing archived data and flows of contemporaneous data streams; and transaction-based, opinion-related and temporally-changing data. We also need to conduct more research to understand the entire life cycle involving data in organizations, including the highly digital environments of e-commerce firms, and also the operations of other firms in more traditional industries. Our emphasis should be on understanding the processes of data generation, data acquisition, data transformation and data integration. We also need to consider data cost and quality, data retention, data analytics and visualization, and data infrastructure and governance. In a word, industry needs many more innovative ways to manage and use big data to support “smart” e-commerce.

With these challenges in mind – and with the added impetus of the flood of big data news flowing from the many business press sources, in 2008 we initiated discussions about constructing a special issue on the topic of “Business and Data Analytics”. For some years already, understanding the outcomes based on working with large data sets was a part of our professional agenda – at the University of Minnesota and Arizona State University, and also at IBM Information Management Group, IBM Global Services and IBM Research. Our work touched on these issues in a number of different industries, including financial services and financial markets, book-selling, telecommunications, IT services and cloud computing, Internet gaming, digital entertainment, travel and hospitality, the airlines, corporate data management and data warehousing – and also many forms of Internet-based selling and electronic market operations. In addition, we represent somewhat different disciplinary perspectives, including Computer Science, Economics, Information Systems, Marketing and Services Science. Taken together, these things gave us an early appreciation of the issue space associated with big data and the opportunity to make a meaningful contribution of new knowledge. This is the background of the present special issue.

We are pleased to showcase new articles on business and data analytics from authors who have some interesting ideas, methods and insights to offer. It clearly is impossible to cover the full range of issues related to the special issue’s theme in a single volume, and hence no attempt is made to do so. We hope that ECRA’s readers will find the coverage offered to be valuable for stimulating new ideas in research and innovative managerial applications. The applied topics in the special issue range from decision support and pricing in the airline industry to understanding how to mitigate the effects of anomalous under-pricing in e-procurement. They also include the analysis of new service adoption for current customers of 3G telecommunications services, and the assessment of network effects in “Business 2.0” e-commerce operations. There are also several useful contributions in the area of mechanism design, an important area for business and data analytics in e-commerce. They touch on issues related to the identification of shill bids in online auctions, the circular relationship between product popularity and rankings in Internet-based selling, the design of personalization approaches, and the refinement of statistics and machine learning methods for identifying relevant data clusters in business operations. Also offered are thought-provoking works that discuss the value-maximizing selection of e-commerce technologies based on “gold mining” methods, and the transformation of operational data in e-services through business analytics so they become meaningful tools for business management.

One of the big challenges that managers have to address when large data sets are available is how to determine what is relevant and then to use it to benefit the business. In the lead-off article of the special issue, Li-Chen Cheng and Li-Min Sun use fifty million data points on subscription customers from the largest 3G mobile services provider in Taiwan, as a basis for understanding their purchasing behavior in that marketplace. The authors observe that now is an especially challenging time for 3G mobile services firms, as their high margins erode in the face of new market competition,
as new products services flow into the market. The authors develop methods for the analysis of customers and how to place them in appropriate segments. They also assess their usage behavior across clusters based on the different kinds of services they consume, and their characteristics as low-value, medium-value and high-value customers. They also develop rules to distinguish between customers in the different value segments, and map out a set of related dependency networks. The dependency networks are useful for understanding the relationship between voice, data and message services, as well micropayments and entertainment goods. This work is important for the techniques that it applies involving experimental design and controls, and the analysis of big data. Thus, we believe it will be of value to researchers and managers interested in implementing experimental research methods in joint academic-industry research.

More than ever before, customers expect personalization from the e-commerce applications that support the services they receive. The second paper of the special issue is entitled “How to Design Personalization in a Context of Customer Retention: Who Personalizes What and to What Extent?” The authors, Kwiseok Kwon and Cookhwan Kim, define a method for designing personalization systems that involve four key constructs: (1) the object of personalization, (2) the personalization level that is provided, (3) the subject of personalization, whether user or customer or firm-initiated, and (4) the preferred learning method for personalization of the firm, whether involving the customer or involving the firm or a marketing research company. The authors conducted a survey of users of the Korean Internet portal, Naver.com, as a means of discovering the efficacy of personalization approach choices on customer retention. Their main findings suggest that interface personalization is important, one-to-one personalization may be less beneficial than is widely believed, and firms need to consider the incentives that ought to be applied in order to increase customer commitment and self-initiation in the process of personalization.

Improving the performance of prediction and estimation techniques for use in e-commerce technology selection is an important problem for research and business. Ohbyung Kwon, in “A New Ensemble Method for Gold Mining Problems: Predicting Technology Transfer,” begins with the observation that conventional ensemble methods typically outperform methods that use single classifiers. The author studies what he calls “gold mining” problems in e-commerce, which involve the identification of the appropriate choices to make among e-commerce technologies that might be worthwhile to patent and invest in, as a basis for doing technology transfer from conceptualization to application in business. Kwon suggests that choosing the “right” e-commerce technology to transfer is similar to finding gold in the mountains: there is a low frequency of a desirable outcome – choosing technologies that generate high payoffs – and a high frequency of an undesirable outcome – choosing technologies that never really pay out. The author develops a “vertical boosting with rewarded vote strategy” as a means to make appropriate predictions. The author’s approach generates classifiers for each of the attributes in a sample. He also provides a way to create rules that are tuned based on their level of sensitivity in order to predict the desirable outcomes. The rule sets that are created can be adjusted further, as a means of generating combined rules that characterize the proposed ensemble method.

The challenge of managing e-services such as e-banking, e-procurement, e-government and e-learning, as well as using the Internet to deliver them, has been increasing during the past ten years, as we witness the continuing global transition from manufacturing economies to service economies. The next article is “Transforming E-Services Evaluation Data into Business Analytics Using Value Economies to Service Economies. The next article is "Transforming E-Services Evaluation Data into Business Analytics Using Value Economies to Service Economies." The authors leverage well-known theory and methods from IS research, including the technology acceptance model and the IS success model, and develop a three-layer value model that covers the multiple concerns of efficiency, effectiveness and impact on users’ future behavior. They also produce useful business analytics, including average user ratings for the strengths and weaknesses of the e-service, and how first-layer value drivers affect others at the second and third layers. Thus, the authors’ research yields a blend of theory-based empirical analysis, as well as new ideas for how to assess the value of e-services.

Today there are many Web applications in the market whose success depends on the number of people that utilize the services they offer. Electronic auctions require numerous buyers and sellers to be of value to their participants, similar to recommender system Web sites and social networks. The next article, by Charu Aggarwal and Philip Yu, “On the Network Effect in Web 2.0 Applications”, is an interesting business analytics methods piece, whose approach can be broadly applied to the adoption of a range of e-commerce applications in which the number of participants matter. The authors apply a Markov chain model to study the dynamic probabilistic process by which Internet auction participants join and leave auctions and social networks, as a means to understand the steady-state behavior of the network of participants that are involved. The research conducted by the authors offers a number of key observations and conjectures that are analyzed with the appropriate mathematics to lay out constructs such as minimum critical mass, participant defection functions, and time-averaged state probabilities. The authors demonstrate the efficacy of using experimental simulation as a basis for assessing the robustness of their Markov model approach. This research work is notable for its interpretation of the underlying network effect conditions that permit an electronic auction or social networking Web site to develop a large following, and then to even out-compete other existing players in the same marketplace.

Clustering is a popular and important data mining technique for the analysis of e-commerce data. In the next article of the special issue, authors John Carls and Kelsey Bruson attempt to shed light on the following research question: In a large data set that is being analyzed, what can we do to go beyond guessing how many clusters should be reported to characterize the empirical data? Their article is entitled “RSQRT: An Heuristic for Estimating the Number of Clusters to Report”. It offers a comparison between an existing approach, the Bayesian information criterion (BIC), and the proposed new approach. The new approach’s acronym “RSQRT” stands for “recursive square root”. It provides a recursive square root heuristic for cluster analysis for data sets that have some number of items (n) or attributes (t) with different scales (e.g., binary, category, continuous), and for which some number of clusters (k) needs to be identified. So, for a data set of 250,000 items (n), RSQRT(n) is RSQRT(250,000), which yields successive square roots until the outcome value is less than 2.25. The process produces a set of four roots: {500, 22.36, 4.73, 2.17}. This suggests that four clusters are appropriate, since the fourth root of 2.17 is less than 2.25. This approach is based on the authors’ argument that square root recursion is the simplest non-linear reducer, and that it works better than other candidate roots based on empirical assessments. The authors applied their approach to data sets so that articles from various published sources could be assigned to an appropriate set of clusters. The authors’ approach performs better in terms of computational cost for identifying the appropriate number of clusters in comparison to the BIC-based approach.
Many e-commerce firms provide capabilities to assist and influence the consumer purchase decisions. This is especially true in the travel and hospitality industries, where there are so many alternative choices for consumers, including airline carrier, route and price combinations, and hotel property, loyalty program and price combinations. The next paper in this special issue, by Eran Rubin and Benny Mantin, is about “Tabulated Decision Aids and Airfare Pricing”. The authors ask: Do online decision support tools that assist consumers in their search for travel date and price combinations result in different pricing approaches on the part of the airlines? Are consumer decisions influenced by the newly-available tools and prices? They investigate the economic effects of decision support tools on the structure of airline prices in the market. The authors’ hypotheses involve: the number of available travel dates and the variance and average of the lowest prices for airfares for different purchase dates. They used one-way fares obtained from ITA Software (www.ita.com), a subsidiary of Google, for 54 US origin—destination city-pairs, as a basis for conducting regression to evaluate their hypotheses. Rubin and Mantin report that more information reach lessens the variability of prices over time, and also leads to fewer price changes and higher average prices.

In electronic market operations, the quality of the mechanism design and the trustworthiness of the marketplace are critical considerations in the public’s assessment of their value. When markets have problems – for example, inappropriate market-maker fees and dealer commissions, fraudulent bidding and insider trading, and operational downtime and trade settlement failures – potential participants become wary of whether a given means of digital intermediation is of value. This often leads to less-than-best participation, and limits to the value of the market. It also sets up the possibility of future failure of the e-marketplace. Sol Shatz, Fei Dong, Haiping Xu and Dibyen Majumdar focus on the problem of fraudulent bidding, in their article “Price Comparison: A Reliable Approach to Identifying Shill Bidding in Online Auctions”. Shill bidding occurs when an auction has bidding participants who are the confederates of the seller. They make bids on sale items that are intended to stir up interest among honest bidders and support higher final transaction prices. The authors study shill bidding in terms of its likely effects on the final prices of sale items sold on eBay. Their approach involves hypothesis testing to determine whether there is a recognizable difference between expected and final auction prices. They use neural network analysis to establish the values to predict final auction prices, with intrinsic data that describe the sale items, and other data related to the auction environment of eBay. They further use a chi-square test of independence and logistic regression to produce a set of experimental results on using lower-than-expected and higher-than-expected prices as a basis for detecting shill bidding behavior.

It is common for managers of e-commerce businesses to ask relatively simple questions about their firms’ activities, but often the answers they hope to obtain cannot be easily acquired. Such is the case for the next research article that is presented in this special issue. Byungyoon Yoo and Kwansoo Kim, the co-authors, ask: “Does Popularity Decide Rankings or Do Rankings Decide Popularity? An Investigation of Ranking Mechanism Design”. This is an important question for research since it pertains to the mechanism designs and services of many digital intermediaries. In this research, the authors explore the business problem that an online music distributor faces in developing ranking policies to maximize its own value to its users. They posit that a ranking mechanism is a communication system between the intermediary and online users, and that the provider can leverage subscription and item download fees, while managing item popularity through the bandwagon, ranking and slot effects. In this research, the authors study online music downloading behavior. The “slot effect” they refer to is the diminished probably that a service subscriber will listen to lower-ranked songs that take more effort to find, since the subscriber needs to work harder to find them further down the list, and there is an exponential degree of attention decay. The authors’ ranking application considers a song’s ranking, and its streaming and downloaded volumes, which are used to assign different value weights. Their ranking application also considers an 80/20 Pareto distribution-based rule for identifying how the slot effect might be related to high and low rankings. The authors’ empirical findings validate a number of proposed hypotheses that describe the link between ranking and popularity, the effects of subscription fees and item unit prices, and the slot effect. In the special issue context, this research is especially valuable in demonstrating how much can be learned from a single month of music ranking chart data. The opportunity to learn much more from several years worth of data would make an interesting doctoral dissertation.

The penultimate paper of the special issue is titled “A Rank-and-Compare Algorithm to Detect Abnormally Low Bids in Procurement Auctions”. The authors are Per Luigi Conti, Livia De Giovanni and Maurizio Naldi. They extend the application of algorithm-based methods in business and data analytics to the detection of procurement auction bids by suppliers who bid very low prices and then are unable to provide the follow-up services or goods that reflect what the buyer believes was purchased. The authors’ approach is to create a rank-and-compare algorithm that helps to identify bidding suppliers whose bids are inappropriately low. These create significant risks for the buyer should the bidder win the auction. This also empowers the auction operator to establish information about the probability of detecting such problems, and gauging whether there might be a mistaken identification of a supplier who is alleged to be unable to perform, but who actually could deliver as expected. The authors’ approach is tuned to deal with fewer than twenty bidding suppliers and bids that are limited in their variance. According to the authors, more bids and a large range of bid prices will make it harder for the algorithm to successfully detect anomalous bids. Such analytics are useful for the management and control of e-procurement auctions, and suggest the variety of ways that buyers can benefit from innovative services that reduce their risks.

The articles that we have included in this special issue are representative of the variety of interesting research efforts that are underway among interdisciplinary researchers and computational social scientists. We took a “developmental reviewing” approach to the various submissions that we received. Some were rejected on the first round of review due to lack of relevance to the main themes of the special issue. When they were of sufficient quality otherwise, we shared the articles for review at Electronic Commerce Research and Applications as Regular Research papers to be handled by the usual review process. For those articles that we did take forward, our emphasis with the reviewing feedback for the authors was to help them find their own voices for innovation and the creation of new knowledge in the business and data analytics domain. We typically used two or three knowledgeable reviewers, although the Guest Editors team also pitched in by writing reviews for some of the articles. Most of the articles received at least two but mostly three rounds of reviews, prior to a final editing pass before publication.

We encourage other authors to send their papers related to business and data analytics for review and publication at Electronic Commerce Research and Applications. The journal is open to works that involve this general theme in interdisciplinary research, Computational Social Science, and a variety of disciplines that are data-focused, including Statistics, Machine Learning and Computer Science. We also welcome innovative works that apply analytical models from Economics as a basis for identifying normative guidelines for consumer, firm and market behavior, and then use large
data streams to test the findings. We further encourage other approaches that blend experimental and simulation approaches, along with Internet-based survey instrument methods and judgment tasks, in the repertoire of classical research methods in social science.

We would like to thank a number of people who assisted this effort in various ways during the three years that we spent developing this special issue. They include past Electronic Commerce Research and Applications Editorial Assistant, Huang-chi Lin, and the current assistants, Annie Yang for Regular Research submissions, and Tong Li for Special Issue submissions. A couple of the papers that we've included in this special issue were initially submitted as Regular Research papers, but due to their aligned thematic content, we decided that it would be appropriate to include them here. We further acknowledge the journal’s Co-Editors, Chris Westland and Kai-Lung Hui, as well as the journal’s Area Editors, Lihui Lin, Nelson Granados and Chuck Wood, for their help in this respect. We also appreciated the assistance of Geetha Krishnan, Mavis Li and Rebecca Wilson at Elsevier, who have given us outstanding support through the special issue development process. Finally, Rob Kauffman would like to thank colleagues Steve Miller, Ee-Peng Lim, Hoong Chuin Lau, Archan Misra and Steve Fienberg, at the Living Analytics Research Center (LARC), a joint venture of Singapore Management University and Carnegie Mellon University, for their encouragement and interest in interdisciplinary research on business and data analytics.

Robert J. Kauffman
Singapore Management University, Singapore
Dartmouth College, United States

Jaideep Srivastava
University of Minnesota, United States

Jamshid Vayghan *
IBM Enterprise Transformation, United States

* Corresponding author.
E-mail address: vayghan@us.ibm.com

Available online 20 January 2012