International Journal of Computer Science and Applications,

Vol. 11, No. 3, pp. 116 – 127, 2014

ANALYTICS ON BIG AVIATION DATA: TURNING DATA INTO INSIGHTS

RAJENDRA AKERKAR

Western Norway Research Institute
Sogndal, Norway
rak@vestforsk.no

The business world is undergoing a revolution driven by the use of data and analytics to guide decision-making. While many forces are at work, a major reason for the business analytics revolution is the rapid proliferation of the amount of data available to be analysed. Recent days, big data is beginning to have a major impact on air travel with more data being created both through the plane sensors and the passengers on board; the opportunities to use this data will only increase. It provides innovative companies with the opportunity to improve major aspects of their business, from using data to improve customer retention through to making planes safer and more reliable. In this paper we discuss a big data concept, definitions, and further present some cases for aviation industry to analyse data from every conceivable channel, for instance, customer data to create a unique profile for each customer based on a wide range of demographic data, behaviours, and preferences.

Keywords: big data; data processing; aviation; analytics.

1. Big Data Concept

Data, or pieces of information, have been collected and used right through history. However, in contemporary world, advances in digital technology have considerably boosted our ability to collect, store, and analyse data. All of the data, however, are merely that—data—until they are analysed and used to inform decision-making.

The use of the term "big data" can be traced back to debates of managing large amount of datasets in both academia and industry during the 1980s. Big data arose due to the emergence of three major trends. First, it has become economical to generate a broad kind of data, due to inexpensive storage, sensors, smart devices, social software, multiplayer games, and the Internet of Things. Second, it has become inexpensive to process huge amounts of data, due to progresses in multicore CPUs, solid state storage, cloud computing, and open source software. Thirdly, not just database administrators and developers, but many more people (such as decision makers, domain scientists, application users, journalists, and ordinary consumers) have become involved in the process of generating, processing, and consuming data. This is known as a democratization of data.

As a result of these accelerating trends, there is now a widespread realization that an unprecedented volume of data can be captured, stored, and processed, and that the knowledge gleaned from such data can benefit everyone: businesses, governments, academic disciplines, engineering, communities, and individuals.

With varied data provisions, such as sensor networks, telescopes, scientific experiments, and high throughput instruments, the datasets increase at exponential rate [Szalay et al 06, Lynch 08]. The off-the-shelf techniques and technologies that are available to store and analyse data cannot work efficiently and adequately. The technical challenges arise from data acquisition and data curation to data analysis and data visualization.

Big data has changed the way that we adopt in doing businesses, managements and explorations. Data-intensive computing is coming into the world that aims to provide the tools that we need to handle the big data problems. Data-intensive science [Bell et al 09] is emerging as the fourth scientific paradigm in terms of the previous three, namely empirical science, theoretical science and computational science. Long ago, researchers describing the natural phenomenon only based on human empirical evidences, so we call the science at that time as empirical science. It is also the beginning of science and classified as the first paradigm. Then, theoretical science emerged some hundreds years ago as the second paradigm. However, in terms of many complex phenomenon and problems, scientists have to turn to scientific simulations, since theoretical analysis is highly complicated and sometimes inaccessible and infeasible. Subsequently, the third science paradigm was the computational one. Simulations usually generate a large volume of data from the experimental science, at the same time; increasingly large data sets are created in various pipelines. There is no doubt that the world of science has changed just because of the increasing data-intensive applications.

In 2012, Gartner recorded the ‘‘Top 10 Strategic Technology Trends For 2013’’ [Savitz 12a] and ‘‘Top 10 Critical Tech Trends for the Next Five Years’’ [Savitz 12b], and big data is listed in the both places. Without doubt, in near future big data will transform many fields, including business, the scientific research, and public administration.

In order to discuss about various issues related to big data, it is necessary to understand different facets about big data. For the definition of the big data, there are various different explanations from 3Vs to 4Vs. Doug Laney used volume, velocity and variety, known as 3Vs [Laney 01], to present data related challenges. In literature, we come across definitions of big data using these 3 Vs:

- Volume – data sizes will range from terabytes to zettabytes.
- Variety – data comes in many formats from structured data, organized according to some structures like the data record, to unstructured data, like image, sounds, and videos which are much more difficult to search and analyse.
- Velocity – in several applications, like smart cities and smart planet, data continuously arrives at possible very high frequencies, resulting in continuous high-speed data streams. It is crucial that the time needed to act on such data be very small.

Occasionally, people use another V according to their special requirements. The fourth V can be value, variability, veracity, or virtual [Zikopoulos et al 11]. In general, big data is a collection of massive data sets with an immense diversity of types so that it becomes difficult to process by using state-of-the-art data processing approaches or traditional data processing platforms.

1.1. *Common Definitions*

Sub-headings Big data is still in its early stages, everyone is still trying to grasp its core nature and to define it scientifically and pragmatically. Nonetheless the precise meaning of the concept remains unclear and is often used synonymously with other related concepts such as Business Intelligence (BI) and data mining[Akerkar & Lingras, 08]. Several stakeholders have created new definitions or revisions of existing definitions that best suit their interests. Nevertheless, to capture the core of big data, consistent themes can be found by examining various definitions provided by the industry gurus and related literature.

Among several definitions reported in the literature, the first formal, academic definition appears in a paper submitted in July 2000 by Francis Diebold in his work of econometrics and statistics [Diebold 00]:

"Big data refers to the explosion in the quantity (and sometimes, quality) of available and potentially relevant data, largely the result of recent and unprecedented advancements in data recording and storage technology. In this new and exciting world, sample sizes are no longer fruitfully measured in "number of observations," but rather in, say, megabytes. Even data accruing at the rate of several gigabytes per day are not uncommon."

The most popular definition in recent years uses the "Three V's": volume (size of datasets and storage), velocity (speed of incoming data), and variety (data types). Reacting fast enough to deal with data velocity is a challenge for most organizations. As mentioned in the section 2, the concept was first raised by Doug Laney (2001) in his META Group research note that describes the characteristics of datasets that cannot be handled by traditional data management tools. With increasing interest and insight in the field, the "Three V's" have been expanded to "Five V's": volume, velocity, variety, veracity (integrity of data), value (usefulness of data) and complexity (degree of interconnection among data structures) [Armour 12]. However, the soul of these V's remains within the extent of data characteristics per se.

Another noteworthy definition presented by Brian Hopkins and Boris Evelson. According to them, if we simply have high volume or velocity, then big data may not be appropriate. The definition described in "Expand your digital horizon with big data"[1]:

"Big data: techniques and technologies that make handling data at extreme scale economical."

The two main characteristics are volume and velocity, while variety and variability shift the curve. In other words, extreme scale is more economical, and more economical means more people do it, leading to more solutions.

More comprehensive definitions and descriptions have also emerged.

For example,

In its report, "Demystifying big data", the big data commission at the TechAmerica Foundation offers the following definition:

[1]http://www.asterdata.com/newsletter-images/30-04-2012/resources/Forrester_Expand_Your_Digital_Horiz.pdf

"Big data is a term that describes large volumes of high-velocity, complex, and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information" [TechAmerica 12].

Researchers at McKinsey propose a *subjective* definition:

"Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyse" [McKinsey 11].

The definition of big data can vary by sector, depending on what kinds of software tools are normally available and what sizes of datasets are common in a particular industry. Big data in several sectors at the moment range from a few dozen terabytes to multiple petabytes.

IBM[2] states that big data involves notable amounts of data that comes from a wide variety of sources. IBM highlights the increasing speed of data generation. "Every day, we create 2.5 quintillion bytes of data — so much that 90% of the data in the world today has been created in the last two years alone. This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few. This data is big data."

"Big data is the frontier of a firm's ability to store, process, and access (SPA) all the data it needs to operate effectively, make decisions, reduce risks, and serve customers" [Gualtieri 12].

A mathematically sound definition of big data:

"Big data represents the historical debri (observed data) resulting from the interaction of at between 70 and 77 independent variable/subjects, from which non-random samples of unknown populations, shifting in composition with a targeted time frame, can be taken" [Smith 12].

Some more definitions of big data presented in literature are as follows:

Association for Data-driven Marketing & Advertising (ADMA)[3], Sydney has provided global, commercial and marketing definitions of big data.

Global definition of big data

Big data is the collection of large volumes of varied information, used to extend our understanding of the environment, medicine, science, business and human experience.

Commercial definition of big data

Big data is the current term given to the wide use of data collected from digital, technological, and analogue sources. Big data is used to improve business understanding of markets, allowing improvements in customer experience and organisational performance.

Marketing definition of big data

Big data is the current term given to collecting, analysing and generating insights from a wide variety of customer, commercial and environmental information.

[2] http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

[3] http://www.adma.com.au/

It is used to develop better understanding of customer preferences, habits and considerations in making transactions with different categories, brands and channels.

The successful use of data in marketing leads to improved customer experience, a better exchange of value between customers and organisations, and improved business performance.

SAP[4] offers a more promotion oriented view on big data.

"Big data is an opportunity to change how you work, play, and live by tracking new signals within digital noise. From major league sports and personalized shopping to carbon reduction and cancer treatment, organizations are using big data to re-imagine achieving what is possible."

Clearly, SAP is focusing on the benefits of big data rather than delivering a straightforward definition of the concept. Besides, SAP is underlining the fact that big data can provide value in a wide variety of fields.

SAS[5] has added two additional dimensions to complement the original three Vs. These are variability and complexity. IBM, includes a fourth V; veracity. Furthermore, a fifth V, value is commonly associated with big data.

Principal analyst for O'Reilly Radar, Edd Dumbill, has given alternative definition [Dumbill, 12]: "Big data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the strictures of your database architectures. To gain value from this data, you must choose an alternative way to process it."

In other words, there is no consensus on the exact characteristics of big data. Nevertheless there are multiple characteristics which most of the vendors in the field agree upon.

From these definitions, we can describe big data as:

> *Using big volume, big velocity, big variety data asset to extract value (insight and knowledge), further confirm veracity (quality and trustworthiness) of the original data and the acquired information, that demand cost-effective, novel forms of data and information processing for enhanced insight, decision making, and processes control. Additionally, those demands are supported by new data models and new infrastructure services and tools which are able to procure and process data from a variety of sources and deliver data in a variety of forms to several data and information consumers and devices.*

2. Big Data in Industry

Big data is transforming the industry landscape, as enterprises tap into more and more broad varieties of structured and unstructured data with greater speed and complexity. Business has always wanted to derive insights from information in order to

[4]http://www.sap.com/bin/sapcom/en_us/downloadasset.2014-04-apr-24-19.sap-makes-big-data-real-real-time-real-results-pdf.bypassReg.html

[5] http://www.sas.com/en_us/insights/big-data.html

make better, smarter, real time, fact-based decisions: it is this demand for depth of knowledge that has stimulated the growth of big data tools and platforms.

The leading enterprises are now including big data from both within and outside the enterprise, incorporating structured and unstructured data, machine data, and online and mobile data to supplement their organizational data and offer the basis for historical and forward-looking views.

Enterprises that invest in and effectively derive value from their data will have a clear advantage over their competitors — a performance gap that will continue to grow as more relevant data is produced, emerging technologies and digital channels offer better acquisition and delivery mechanisms, and the technologies that enable faster, easier data analysis continue to develop.

Mostly business environment changes frequently and rapidly. Thus, in such environment, future prediction becomes more vital than the modest visualization of historical or contemporary perspectives. For effective future prediction, data analysis using machine learning and predictive modelling techniques may be applied to enhance and support the organization's business strategy [Akerkar, 13]. The gathering and aggregation of big data, and other information from outside the enterprise, enables the business to develop their own analytic capacity and capability, which for many years has only been available to a few larger organizations. The aviation industry encompasses a huge amount of data, and many airlines and airports cannot manage and process the amount of data they receive, but such data could be used to revolutionise the passenger experience. The vast amount of data produced related to passenger flow, cost reduction and revenue enhancement is too much to handle for most small airline IT departments. Data-driven marketing can provide insights from data in real-time so there can be consistent understanding of passenger behaviours.

From recent Accenture[6] study, big data analytics has become the highest priority for aviation (61%), wind (45%) & manufacturing (42%) companies. The following graphic provides insights into the relative level of importance of big data analytics relative to other priorities in the enterprises interviewed in the study:

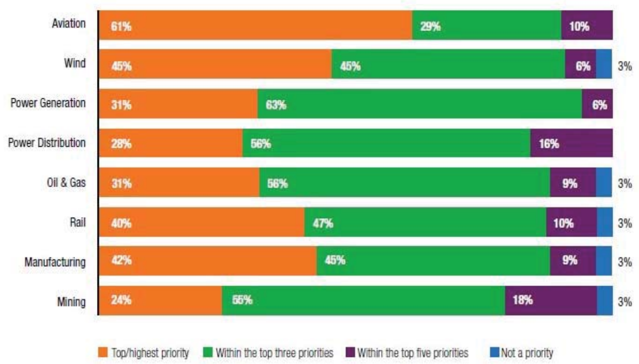

Figure 1 Data analytics priorities in industry[5]

[6] http://www.accenture.com/us-en/Pages/insight-industrial-internet-competitive-landscape-industries.aspx

With 35 million flight departures per year, data is critically important for any planning decision made by airlines and airports. In an example used by Josh Marks, the aviation industry needs to begin documenting and utilising data in a way similar to the online retailer, Amazon.

Amazon manufactures product suggestions based on its customer’s previous activity. They collate data about customer’s previous purchases, as well as previously viewed items, and even mouse movement on the webpage. These predictive analytics allow them to target customer’s needs effectively and with profitability.

This sort of data collection could effectively be utilised in the aviation industry, particularly with airline websites with regards to booking and transactional data sets. There are $140 million worth of ticket transactions made through airline ticketing websites- and this figure is not inclusive of booking sites such as Expedia. The current ‘look to book’ ratio stands at approximately 10:1, signifying that customers will view approximately ten booking websites before making a final decision.

However, airlines do no track this sort of data, only their transactional data, therefore missing out a huge chunk of possible marketing opportunities. Tracking the time and place a customer booked from, fares offered from other websites, and customers IP addresses would allow airlines a greater insight into its customer needs. Probably most importantly, it would allow airlines the opportunity to see if a customer searched for a route it does not offer - this sort of information is fundamental in route development.

It is not just traditional businesses which are harnessing the value of data. Over the last two decades we have seen the rise of a new industry, whose main asset is data. Armed with vast amounts of data and affordable capabilities for storage and processing, companies have been able to create new business and revenue streams from selling data or delivering services that are based on data analytics. The most prominent examples of new use of data occur in the new Internet-based industries. Here, companies exist that have created big business based entirely on access to data and their ability to use data. In addition to innovations in the business domain these companies have also moved “big data technology” forward by inventing and implemented the technology needed to solve the problems related to data handling and analyses in their own operations. Prime examples of such companies are Google, Facebook, Yahoo and Twitter.

As mentioned earlier companies in sectors such as telecom, retail, media, healthcare, insurance, finance and transport/logistics are being transformed by utilizing large data sets about their customers and internal work processes. Insights from these data streams are used for better assessment of utilization, to improve operational efficiency, and to increase sales and the efficacy of marketing.

For traditional producers of goods the main use of data will be to facilitate improvement in the quality of the end product and to make the overall process more cost-effective. Typically this results in an increasingly automated production process. The benefits of process optimization has of course has been recognised for industrialized processes for a long time, but access to new technology, the availability of data have made it possible to take this to take this to new levels of sophistication.

In contrast, businesses in the knowledge industry (consulting would be a prime example of this) the main product is helping customers in solving their problems, and the key assets are the knowledge and capabilities of its workers. The main changes, challenges and opportunities posed by big data in this industry are related to: changes in the competitive landscape, data handling capabilities expected by the customers, and go to market strategies for data driven services. New technical capabilities have the potential to increase the efficiency of project delivery, to facilitate reuse of data, and enable scaling, through automation of the delivered services. At the same time; the possibility of encoding knowledge into models and use these models in combination with data can also enable competition from new players and thus completely alter the competitive landscape.

The advent of more use of sensor data and automated monitoring is the sign of the next wave of changes in business: “The Internet of things” (IoT). A future where an products and appliances not only are able to sense and act upon the environment but also exchange and act upon information and data shared by other devices (i.e. intelligent machine to machine communication). These technologies will certainly take automation to the next level and it will be crucial for future “smart grid solutions” in the power sector, and they will also generate large amounts of machine to machine interaction data which will be available across traditional barriers. The IoT trend will therefore open up new opportunities for advanced analytical offerings and services and this is the main rationale behind the large investments in IoT and analytics by companies like General Electric (GE) and Phillips. Another trend which plays into both commercial use of big data and IoT is the advent of a new generation of cognitive computing, like IBM’s Watson technology, which is able to sift through, combine and infer knowledge form large collections of data sources[7].

Internet of Things has replaced big data to be the most hyped technology as Gartner’s Hype Cycle[8] for Emerging Technologies said. Things here can refer to uniquely identifiable embedded computing devices such as heart monitoring implants, smart thermostat systems, etc. Big data and Internet of Things together will help build valuable systems. The value of IoT will be in gaining timely, valuable insights from all of the data being generated by sensors, etc. The advent of the IoT does also drive development of big data technology, as an example of this ParStream’s analytics database is a great fit for the speed and scale of IoT and is exactly the type of technology that can be the difference between big data and smart data driving business value.

Mostly the discussion around big data has focused on clickstream data, sentiment analysis and consumer targeting. But behind the scenes, the capabilities enabled by machine-to-machine communication and advanced analytics stand poised to dramatically change the world around us. The Industrial Internet of Things (IIoT) is connecting the physical world of sensors, devices and machines with the Internet and, by applying deep

[7] “Disruptive technologies: Advances that will transform life, business, and the global economy”, McKinsey white paper on disruptive technologies, http://www.mckinsey.com/insights/business_technology/disruptive_technologies

[8] http://www.gartner.com/newsroom/id/2819918

analytics through software, is turning massive data into powerful new insight and intelligence. The Industrial Internet involves putting different kinds of sensors, sometimes by the thousands, in machines and the places they work, then remotely monitoring performance to maximize profitability. G.E.[9], one of the world's biggest makers of equipment for power generation, aviation, health care, and oil and gas extraction, has been one of its biggest promoters.

The development in IIoT will allow us to connect intelligent machines, advanced analytics and clever people working in smarter ways to achieve extraordinary things. And the fuel driving this revolution will be big data – masses of information aggregated, analysed and put to work across sectors as diverse as energy, transport, aviation, healthcare, consumer goods, retail and even other professional services. The industrial internet promises great benefits. However it raises some sensitive legal and moral issues, which will become increasingly prominent as the use of data and analytics becomes more sophisticated.

3. Insights through Big Data Analysis

Recent days the aviation industry adopts on Condition/Preventive maintenance procedures due to its operational efficiency and it depends upon the failure mode calculations made after testing a part under circumstances. These conditions may fluctuate depending on the external factors/human errors which may result in the variation in the life time of components in turn reducing the operational efficiency of the aircrafts.

A study by FAA[10] states that during a year jet engine generates data equivalent to 20TB. Basically most of this data is not used for any of the analytics purpose since this data is unstructured. Big data analytics can be used to predict the fault in the component by analysing data obtained from various sensors and account of the specific component attached to an aircraft or fleet in which it is present.

As we know most airlines collect lot of data for historical, archiving and compliance purpose. However, there are very few attempts towards acquiring the customer intelligence from the massive data sources. There is a clear need for drastic shift from conventional analysis to provide insight about the past & current state of business, to next generation systems deployed for predictive analytics of latent customer behaviour.

In order to obtain such insight airline companies must utilise an integrated outlook of customer which covers different touch points and numerous data sources that may or may not reside within the company's ecosystem. For instance customer profiles, transaction details, social behaviour, complaints, psychographic data, techno graphic data etc. They would need cutting-edge big data ecosystem to make sense of such structured & unstructured data sets characterized by large volume, high velocity and great variety.

In the following subsection, let us define some use cases.

[9] http://www.ge-ip.com/industrial-big-data

[10] http://www.aia-aerospace.org/assets/FAA_2013_to_2033_Aerospace_Forecast.pdf

3.1. ***Typical use cases***

1. Assume that a pilot augments their decision making process with real time business intelligence, for instance, information available in the cockpit would allow them to make appropriate tunings to their flight patterns. The computational methods to better optimize flight paths so airlines can reduce cost, avoid bad weather, and get to their destinations on time.
2. Airlines produce enormous amounts of data during various processes such as online communication, online price comparison and ticket purchasing, online check-in and seat selection, personalization of offers, etc. Some airline companies use complex algorithms to collect and analyse the vast amounts of data that are generated in real-time by the sensors nowadays present on every plane. This is tapping into industrial internet. The sensors help airlines to monitor the planes in real-time, also when in flight, and manage and predict maintenance, spot problems before they happen, reduce fuel consumption and shorten turn-around-time at airports.
3. To understand customers better than its competitor, utilise the customer intelligence amassed across several touch-points. Exploit it to perform behavioural marketing, recommend products & services, personalize their offering for every customer in an enhanced manner, up-sell & cross-sell and develop new revenue stream altogether. For example, personalization can be like providing a different look and feel for distinct customers at home page of airlines website.

3.2. ***Generating key insights***

In this subsection we consider a case of an airline market competition.

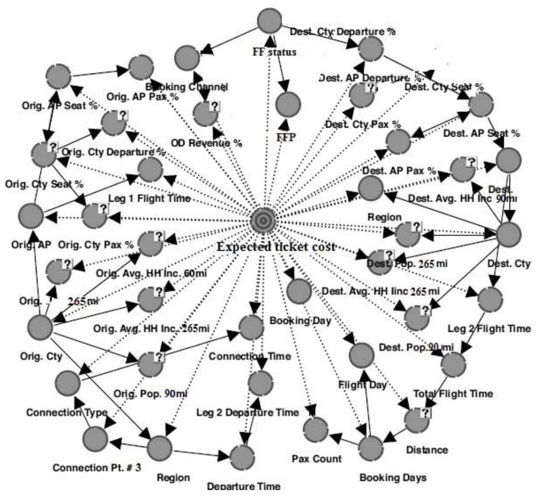

Figure 2 Bayesian network for passenger pattern

Until now most airlines have analysed partial datasets, which are thinned subsets of the big data. Thus the airline companies haven't been able to extract the broad understanding it was looking for. In order to extract important insights from the big data that would not be readily apparent from studying limited localised datasets, we have used tools such as R[11], Tableau[12] and BayesiaLab[13]. A Bayesian network is flexible in dealing with sparse and incomplete datasets. By using Bayesian networks, big data technology and airline market knowledge, we have carried out computation to explore the big data repository and to visualise the results. Bayesian maps are used in supplying the missing information and details as well as bringing the priorities and importance of the effective factors. In comparison to other casual maps, Bayesian maps use Bayes theory to find the certainty factors on the interaction of decision variables. Bayesian network is applicable when historical data is available and shows the major advantage on analysing complex unstable systems. We analysed large datasets containing one year of passenger data – approximately 40 gigabytes. The outcome from mining the data included inferences about airline performance, market traits and passenger booking patterns as shown in Figure 2. We generated hypothesis to test from the data, such as: can predictive modelling be developed to evaluate routes for reservation changes, and to schedule profitability for prospective expansions?
More cases and detailed results will be presented in extended version of this paper.

4. Conclusion

With the advent of the big data era, dealing with large amounts of data is challenging. Big data can help operations for airline companies and airports to reduce redundant variability. The airlines of the future will be defined by the ability to obtain & process the enormous amount of customer data to gain insight at a speed surpassing the pace at which customer behavior and business environment might change. Using the data furnished, airlines can offer a personalized incentive for every type of customer resulting in more auxiliary sales, greater percentages of repeat business, and better customer allegiance. Further it can help the relationship between airports and airline companies in the operational sense, and help build knowledge about demographics.

References

Akerkar, R. A. (2013) Big Data Computing. CRC Press.

Akerkar, R. A.; Lingras, P. (2008). *An Intelligent Web: Theory and Practice*, 1st edn. Johns and Bartlett, Boston.

Armour, F. (2012) Introduction to big data, presentation at the symposium Big Data and Business Analytics: Defining a Framework, Center for IT and Global Economy, Kogod School of Business, American University, Washington, DC.

Bell, G., Hey, T., Szalay, A. (2009) Beyond the data deluge, Science 323 (5919), 1297–1298.

[11] www.**r**-project.org/
[12] www.tableau.com/
[13] www.bayesia.com/en/products/bayesialab.php

Boland, D., Morrison, D., and O'Neill, S. The future of CRM in the airline industry: A new paradigm for customer management. IBM Institute for Business Value. http://www-05.ibm.com/innovation/nl/pdf/highlights/integration/crm_airline.pdf

Diebold, F.X. (2000). Big data dynamic factor models for macroeconomic measurement and forecasting. Discussion read to the 8th World Congress of the Econometric Society, Seattle, August. http://www.upenn.edu/~fdiebold/papers107/ABCD_HOED.pdf.

Dumbill, E. (2012) Making sense of big data. DOI: 10.1089/big.2012.1503.

Gualtieri, M., The pragmatic definition of big data. Forrester Research Blog. http://blogs.forrester.com/mike_gualtieri/12-12-05-the_pragmatic_definition_of_big_data

John-Paul Clarke. (2013) Enabling Pro-Active Decisions via Predictive Analytics. Air Transport IT Summit.

L. Neumeyer, B. Robbins, A. Nair, and A. Kesari. (2010) S4: Distributed Stream Computing Platform, in Proceedings of IEEE International Conference on Data Mining Workshops (ICDMW), pp. 170–177.

Laney, D. (2001) 3d Data management: controlling data volume, velocity and variety, Appl. Delivery Strategies Meta Group (949). http://refhub.elsevier.com/S0020-0255(14)00034-6/h0650

Lynch, C. (2008) Big data: how do your data grow?. Nature 455 (7209), 28–29.

Mark van Rijmenam. (2014) Southwest Airlines Uses Big Data To Deliver Excellent Customer Service. Datafloq blog. https://datafloq.com/read/southwest-airlines-uses-big-data-deliver-excellent/371

McKinsey Global Institute. (May 2011) Big data: The next frontier for innovation, competition, and productivity.

Nathan Marz and James Warren. (2012) Big Data- Principles and best practices of scalable realtime data systems. Manning publishing.

Savitz, E. (2012) Gartner: 10 Critical Tech Trends for the Next Five Years. http://www.forbes.com/sites/ericsavitz/2012/10/22/gartner-10-critical-tech-trends-for-the-next-five-years/.

Savitz, E. (2012) Gartner: Top 10 Strategic Technology Trends for 2013. http://www.forbes.com/sites/ericsavitz/2012/10/23/gartner-top-10-strategic-technology-trends-for-2013/.

Smith, J.A. (2012) Field Note: What Makes Big Data Big – Some Mathematics Behind Its Quantification, Data Scientist Insights. http://datascientistinsights.com/2012/12/19/field-note-what-makes-big-data-big-some-mathematics-behind-its-quantification/

Szalay, A., Gray, J. (2006) Science in an exponential world, Nature 440, 23–24.

TechAmerica Foundation. (2012) Demystifying Big Data. Washington, DC.

Zikopoulos, P. Chris, E. (2011) Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGraw Hill Professional.