LEVERAGING THE POTENTIAL OF BIG DATA ANALYSIS IN AVIATION

Eva Endrizalová[1] - Stanislav Szabo[2] - Rostislav Pšovský[3]

Abstract: *This article is based on a research of the same name, which is focused on the topic of a relatively new and progressive area of large data sets analysis with the aim to discover hidden patterns and connections in order to support business models of airlines. The first part of the paper justifies the need of using Big Data analysis by describing the unprecedented growth of digital data during last decades. The author also describes the phenomenon of Big Data itself, its advantages over traditional inferential statistics methods and technology demands of large data sets analysis. The second part analyses the possible use of Big Data analysis in the aviation industry. Three different applications are described in detail: Airlines' customers profiling using new data sources and providing them with individual offers, improvement of aircraft predictive maintenance through sensor data anomalous behaviour detection and decision making support at operations control center of the company Travel Service.*

Keywords: Big data, information revolution, data growth, technologies, analysis, correlation, personalized marketing, customer profiling, predictive maintenance, decision making support.

1 INTRODUCTION

We live in the middle of an information revolution. The unprecedented development of information technologies results in generating of massive data volumes in almost any area of human activity. These massive data sets which have developed from gigabytes of data more than 20 years ago into petabytes and even exabytes today due to rapid development of mobile technologies and emergence of data from social media, are often referred to as "Big Data".

Lately, the concept of Big Data has been passionately discussed. The continuing growth of computing power and decrease in cost of data warehousing has laid foundations to new analytical possibilities which could result in a new "industrial revolution" based on data. According to some companies, data represent the most valuable commodity of the 20th century. The power of Big Data lies in the recently gained ability to give answers to questions we haven't been able to answer till this day. They offer an alternative to the classical methods of inferential statistics which make conclusions about the whole data set based on a carefully selected data sample from it. To the contrary, Big Data analysis means analysis of the whole data set and discovering of hidden patterns and connections in data which couldn't have been seen in the past due to lack of computing power and relevant data itself.

Aviation stands in the front line among industries which have the possibility of leveraging the potential of Big Data within their reach. Undertaking in the aviation industry has always strongly relied on data and all airlines have massive volumes of historical data at their disposal because of that. This data is stored among a number of disconnected storages. The data has to be integrated and applied Big Data analysis to in order to recognize hidden correlations. Thorough understanding of these insights will have a significant impact on airlines' business strategies. Aviation is a low-margin industry which pushes companies to continuously search for means of competitive differentiation and market-share growth. Implementation of Big Data analysis into company processes can result in the desired competitive advantage. It is even possible that Big Data analysis will become a common concept and inability to react to insights deduced from analysis of massive data sets will for many companies result in uncompetitiveness.

The concept of Big Data is not easy to understand. Company CEOs feel Big Data is something that requires immediate attention but at the same time they are unable to imagine what value it may bring to their company. My aim in this article is to assess technology requirements of the Big Data infrastructure and its implementation into current airline systems but mainly to describe possibilities of leveraging Big Data by airlines including sales, maintenance and decision making.

2 DATA EXPLOSION

During the course of human history, we've experienced numerous information revolutions. The first significant information revolution, dated around 3000 BC, is connected to development of the first script [1]. Its significance lies in the ability to make a note of available information. It was, however, impossible to spread information among many people because all the information had to be reproduced by handwriting, which was very time consuming. That was changed around 1439 thanks to the invention of Gutenberg's printing press [2]. Between 1453 and 1503, the same amount of books was printed as had been printed over the whole course of the previous history which means that the amount of information had doubled during only 50 years [1].

The information revolution we live in has been caused by the rapid development of information technologies after 1945 and the birth of internet in 1983 [3]. The companies that first leveraged the potential of computers in the 1970s, 1980 and 1990s were focused mainly on transactional data which were of the biggest value to them. For the first time in history it was possible to track in detail when, where and by whom products are bought [4][5]. Already in 1980s, the amount of transactional data reached the order of gigabytes which meant big demands on data storage [6], since typical storage devices were represented by a magnetic tape with the capacity of 225KB, 5,25 inch floppy disk with the capacity of 1,2MB and a hard drive with the capacity of 5MB [7]. By the end of 1980s, the growing popularity of digital technologies propelled the amount of data into the order of terabytes.

In 1989, the former employee of CERN, Tim Berners-Lee, developed World Wide Web (WWW) and pushed the world into the internet era [3][8]. During 1990s, the amount of internet data reached the order of petabytes [6]. In the light of current development, companies will soon have to deal with exabytes of data and this amount of data will introduce new demands on infrastructure, since current technologies are only able to cope with terabytes or petabytes of data [6]. The dynamics of the development of the amount of data can also be explained by the comparison which says that the amount of stored information is growing 4 times faster than the global economy. Parallel to that the computing power is growing 9 times faster. At the front line of the data deluge stand internet companies, e.g. Google, where they analyse more than 24 petabytes of data every day [1].

The internet era is in important aspects similar to the printing press era. Looking at the deep impact of the printing press on its era, we can expect similar impact of internet on this era. Both these technologies represent a major breakthrough in the way how people communicate, how they store, update and widen their knowledge but also in defining of knowledge ownership and how it's obtained [2].

The last decade can be without exaggeration called the decade of mobile revolution which started in 2003 after introduction of the third generation commercial mobile networks. Mobile 3G technology differs from the previous generations of mobile networks by enabling people to use broadband network which made it possible to connect to the internet from any device supporting this technology and placed within sufficient mobile signal coverage [9].

The impact of the mobile networks of the third generation can be demonstrated on the exponential growth of the number of devices connected to the internet immediately after the introduction of their commercial usage. In 2003, there were about 6.8 billion people on Earth while the number of devices connected to the internet was assumed to be 500 million. In 2010, there were already 12.5 billion devices connected to the internet (i.e. approximately 1.84 devices per person) and this figure is still increasing. It is estimated that the number of these devices exceeded 7 billion somewhere between 2008 and 2009. This important event is very often referred to as the emergence of the "Internet of Things" [10]. Estimates of the future development in number of devices connect to the internet vary significantly between 25 and 75 billion in 2020 [10][11][12].

It is almost unbelievable how dependent on computers we have become. If we look back, 150 years ago, most people produced a tiny amount of data. For some these data were limited to the date of their birth, the names of their parents, the place and date of their wedding and the date when they died [13]. Today, people produce a massive amount of social data and willingly or unwillingly share their personal information with third parties which can analyse them and create a complex, almost perfect picture of us. Any company able to recognize a business opportunity in this data can easily turn the knowledge of their customers into profit.

3 BIG DATA

The situation connected with the current explosive growth of data and the potential of their possible use by companies is in today's literature very often represented by the term "Big Data". It is, however, not at all easy to find a unified definition of Big Data. Big Data can be described as the process of complex data set analysis in order to discover hidden information which could support decision making or discover patterns and correlations which have not been previously known [14].

The term Big Data is in fact imprecise. It suggests that the sheer volume of data is the basic problem which has to be solved. The biggest challenge, however, comes from the variety of data. According to some estimates, only 5 % of all digital data is structured, i.e. it can be stored by the means of rows and columns in a relational database [1]. Such data can be imagined as data stored in MS Excel table. If we only were dealing with exabytes of structured data the problem of volume could be easily solved by applying enough computing power. Today's data comes from a wide variety of sources thus differs in its structure significantly. We can imagine it to be text file, an audio recording, a video, pictures, sensor data, clickstream data, GPS data, DNA analysis data, etc. They are massive data sets with low value density by the analysis of which we can reach a deeper understanding of a specific problem [6].

Modern technologies around Big Data represent a hot topic and in most cases are considered the definition of Big Data itself. This definition, however, omits the important fact that Big Data represent mainly a transformation of thinking. Computer process the data but that is not revolutionary itself. Revolution is happening within the data and in the way we work with it. When we have access to massive data sets we are able to run operations we couldn't do with small amounts of data. With the increase in volume comes a change in nature. This opens way to predictive systems relying on big data sets. The outputs of these systems are forecasts which lie in the very core of Big Data analysis [1].

2.1 Whole population analysis

The first obstacle preventing us from unleashing the potential of Big Data lies in the fact that our way of thinking comes from the world of small data, i.e. from the world of data scarcity [15]. In the world of small data, we use methods of inferential statistics which are tools helping us to make a conclusion about the whole population based on a randomly selected sample from it.

Random sampling is a relic from the time of insufficient computing power. Today, when we have enough computing power, storage capacity and cutting edge tools at our disposal, we can mostly replace random sampling by Big Data analysis, which eliminates most of its drawbacks. Using all available data instead of just a little part enables us to observe details and connections which get otherwise lost in the data deluge. We can go through the data in more depth a repeatedly analyse it in a completely new way we haven't even considered while the data was being collected. Big Data enable us to examine new hypotheses at many levels of detail because there are bases on the biggest possible volume of information and the analysis is not endangered by the risk of distortion. When we use all data we can spot even the smallest details which we otherwise wouldn't have been able to observe because they would simply not even be present in the data sample [1].

2.2 From accuracy to approximity

The next field where Big Data make us reassess our current approaches is the accuracy of processed data. In the world of statistical analysis where we try to make conclusions about the whole population based on a tiny sample, the numbers used for analysis have a great impact on the result of the analysis and have to be very accurate. When we analyse only a limited number of data points, the final mistake gets the more imprecise the less accurate data we use. When we use only a tiny amount of data we logically aim to reach its highest possible quality.

Collecting of big data sets gives way to inaccuracies. There are in fact so many in the volume of data that it becomes impossible to get rid of all of them. However, we have to compromise – if we admit inaccuracies in data, we can collect a much bigger volume of it. At the end of the day, we gain more than we lose.

The data cannot be completely faulty, of course. But if are only trying to assess a general trend in data, we can lower our demand on data quality and thus be able to work with a much bigger data set. During data processing, we get rather probabilities than precise values. It is possible to increase accuracy by various means but when we are only trying to assess a general trend it is usually not even economical. Obsession with data accuracy comes from the time when each point of the data set had a significant impact on the overall accuracy of the outcome of an analysis. Today, we trade accuracy for the ability to create a much richer picture about a situation [1].

2.3 Causation vs. correlation

It is necessary to understand that while searching for hidden patterns and connection in data we're not trying to prove causation, i.e. we can't surely say that A was caused by B. When we see that A is somehow tied to B, we say that A and B are correlated. In other words, we use correlation to quantify statistical relationship between two data values. When two phenomena are strongly correlated it means that when one of them is changed the second will change as well with a high degree of probability [15].

Predictions based on correlations lie in the core of Big Data. Data of consumer marketing have for example been identified as a proxy to blood and urine samples analysis, late invoice payments as an indication. For example consumer marketing data has been identified as a good proxy to blood and urine sample analysis, late paid invoice as an indication of a divorce or an interest in healthy lifestyle and buying of unscented body cream as an indication of pregnancy. As well causation, correlation can't be surely proved as. It can only be confirmed with a high degree of probability. One of the advantages of correlations is that they can be expressed mathematically. That is, however, impossible in the case of causation. Correlation analysis will not give us answer to WHY some phenomena relate but in the most of the cases during decision making it is sufficient to know they do relate. In reality, we are not so much interested in the WHY, the WHAT is all we need to know [1].

2.4 Unleashing the hidden value

The immediate value of data collected by companies is usually obvious. The reasons for collecting data by companies are called "primary applications" of data. In other words, it is the purpose for which the data is collected.

The value of data can be compared to an iceberg floating in an ocean. The primary application of data is represented by it tip above water which is impossible to overlook. The potential of data is, however, represented by the rest of the iceberg hidden below the surface and we will only see it if we'll assume it to be there. This potential represents the secondary application of data usage. The more different ways of using one particular data set we can find, the more it will grow in value. Secondary applications require especially an open mind, innovative approach to data processing and a cutting edge set of methods and tools which are being delivered by the new generation of statisticians. In the past, data was usually deleted or archived it served its primary purpose. Today, we already know that it still contains potential value which can be unleashed by certain means [1].

4 BIG DATA TECHNOLOGY

The value chain of Big Data analysis comprises 4 basic phases: *Data generation, Data acquisition, Data storage and Data analysis.*

Data generation is the phase is the phase during which data is created. As noted earlier, the term Big Data is used to indicate a massive, variable and complex data set generated from different distributed data sources, e.g. sensor data, videos, clickstream data and other available sources of digital data.

Data acquisition refers to the process of information obtaining and is further subdivided into data collection, transmission and pre-processing. Data collection is represented by technology used for data collection, through which gross data is obtained from the specific locations of their production. After collection of gross data we need a high-speed transmission mechanism to transmit data into a respective storage relevant to the specific analytical application. Collected data sets can in fact include big amount of nonsensical data which requires additional storage and influences analysis. For example, redundancy is a typical phenomenon during sensor data collection and we try to minimize it via data pre-processing.

Data storage means permanent storage and management of massive data sets. The storage system can be subdivided into two parts: hardware infrastructure and data management. Hardware infrastructure represents an elastic set of sources of information technology. It should meet the requirement of scalability and dynamic reconfiguration due its use in different types of application environments.

Data management software lies over hardware infrastructure and in order to maintain massive data sets. It comprises a files system, database technologies and a programming model. In the field of Big Data, distributed files systems are used which are represented mostly by Google File System (GFS) and its open-source derivative Hadoop Distributed File System (HDFS). They are both scalable files systems (their capacity can be easily increased by additional commodity hardware) designed for highly distributed data applications [6]. NoSQL databases are slowly becoming common in the field of Big Data. Their biggest advantage over relational databases is support of big volumes of structured, semi-structured and unstructured data, quick iteration, support of object-oriented programming and effective scalable architecture based on commodity hardware. Programming models decrease the performance abyss between relational and NoSQL databases by enabling operations over large data sets [6]. MapReduce is the most commonly used programming model for easy development of applications processing massive volume of data on a big cluster (thousands of computers) of commodity hardware in parallel and by reliable means [16].

Data analysis uses analytical methods or tools for research, transformation and modelling of data in order to extract value. The emerging analytical research can be subdivided into six technology areas: structured data analysis, text mining, multimedia analysis, web analysis, network analysis and mobile analysis [6].

5 BIG DATA IN AVIATION

All companies undertaking in aviation produced a big amount of data in the history and this amount is still growing. According to Air Transport Action Group (ATAG) bases in Geneva, 8.6 million people are transported by 100 thousand of flights every day. That means 35 million flights every year during which trillions of data points are created [17]. Each activity connected to conduction of each particular flight leaved a data trace. In summary, hundreds of terabytes or petabytes of structured transactional data in conventional databases and no less than that amount of unstructured data are created [18]. A big airline produces more than 3 exabytes of data every day and this amount doubles every 3 to 4 years. At every moment, a big airline has to be able to store more than 1 terabyte of data [19]. The Global Distribution System of the company Sabre produces up to 7 terabytes of transactional data every day [20]. The same scenario happened on the part of the airframe manufacturers, too. It is for example estimated that Boeing has around 100 petabytes of data at their disposal and they aim to find correlations within that data in order to provide airlines and passengers with better services [21]. Some aircraft can already produce between 500 and 1000 gigabytes of data on an average flight depending on the level of instrumentation [22].

It is impossible to overrate the transformational potential of Big Data in Aviation. Big Data is probably the biggest opportunity for subjects undertaking in the aviation industry to make use of the changing data structure and its utility maximization. They carry the potential to streamline business and at the same time bring a better experience from travelling.

A. Personalization:

Industries which produce and store big amounts of data have the biggest opportunity to create a relevant business value by aggregation and analysis of this data and a proper reaction to its outcomes. From this point of view, the aviation industry probably has a big opportunity in the area of personalization and sales at their disposal, most probably much bigger than other industries. Airlines sit on terabytes of their own customer data containing details about purchasing habits of passengers, historical data about handling processes, cabin service preferences, movement of passengers within and airport, interactions with Customer Care department and many other categories of data.

The key to success is the ability to match the right products and services to a particular passenger, who will have optimal use of it. In order to accomplish that, it is necessary for airlines to analyse their customer data and apply systems of automated rules to offer personalizes offers and proactive customer service. The concept of personalization means the ability to react to specific requests of every single passenger.

The methods of Customer Lifecycle and Market Segmentation are slowly but surely becoming old relics of the era of insufficient data and technologies need to understand complex customer attitudes, behaviors and habits on an individual scale. It is possible to gain customer loyalty via personalization of offered connections, business partnership, catering, in-flight entertainment, customer service, ancillary services, technology innovations, etc. Airlines which will be able to exceed customer expectations will gain the ability of quick absorption, interpretation, personalization and offering of a consistent product which customers will remember and will be willing to use it and pay for it again in the future [19].

Airlines will also have to face numerous technology challenges. The result of a long term extensive use of information systems by traditional airlines is that Big Data technology architectures will have to coexist with current hardware software and databases because traditional tools and data in them are still needed and will be useful for analysis, and improvements of operations and customer relationship. Pure Big Data technologies are a good solution for newly set up companies and purely internet companies but we can expect hybrid environment in current big companies in the near future which will cause further challenges in the area of architecture cohesion of information technologies and effective functioning of both new and old systems. The former data technology introduced by IBM in the 1960s are unable to support open-source environment for Big Data analysis, e.g. Hadoop, which is a commercially exploitable software network including a distributed files system (HDFS) and a MapReduce platform [18].

Modern transactional and operation systems are slowly replacing traditional systems but airline managers are slowly losing interest in transactional data. Airlines at the front of the market stand over their competitors especially thanks to customer data analysis and knowledge of their customers. The developing amount, structure and other attributes of data customers share with airlines causes problems, though. Management and integration of data from various sources is globally the biggest challenge for all professionals in the field of customer analysis. Airlines have tried to aggregate their data for decades a hoped they would create a unified tool of the ultimate truth about their passengers. Considering the fact that data in many companies are distributed over 20 various storages and at big airlines across 50 or more storages, creating a unified, holistic view of a customer is a highly complex and expensive project. Some companies were successful in creating a single customer data management system but they mostly created just another isolated data silo requiring regular re-calibration which resulted in higher disorganization than would have otherwise existed.

Customer profile is the key element needed for realization of a business strategy with a customer in its middle. Many airlines are unable to identify individual customers unless they are members of their loyalty program. The concept of a customer profile enables this mostly because it helps to create complete profiles of all passengers. A customer profile is ideally a 360° view including an exhaustive list of customer attributed. At the beginning, it enlists descriptive attributes such as contact information, demographic and psychographic segmentation, hobbies and stated preferences. The value of a customer is also important. A holistic and accurate view of a long-term potential value could be assessed by indicators such as total profit, contribution to margin, history of ancillary services purchases and loyalty program information. Additionally, the history of interactions of the customer with the airlines is of significance. From personal contact, across call centers and social media posts, processing of interactions in real time helps maintain information awareness of the current status and customer satisfaction, helping to tailor the next interaction to the situation.

Distribution has always been an area where computers, data and analysis have been used extensively. Therefore, it can be assumed that in the first phase Big Data will have the biggest significance in the field of distribution. Some distribution offers are already somewhat personalized e.g. based on loyalty status or purchase behavior but the majority of offers today responds to customer requirements very remotely [18]. Our goal is not to glut a customer with information but to offer them exactly the information they might find interesting while purchasing a product. Big Data give us a wonderful opportunity at the moment when a customer is looking at our product to find out, what he’s interested in and we should best respond to his purchase requirements.

After identification of customer profiles, we are able to process purchase requirements on a personalized level. A purchase requirement goes into customer system in real time and its parameters go to predictive module. The customer system returns requirement availability while the predictive module determines which products and services may be of interest to the individual customer. These pieces of information are merged in a sales modules and the result is presented to the customer via the distribution channel they are using [23].

B. Predictive maintenance

Predictive maintenance represents such a maintenance control program which is based on monitoring of individual aircraft components during operation, assesses maintenance requirements and enables it to be done exactly before a possible critical failure. The transition from preventive maintenance base on component replacement in fixed time intervals is based on information availability including both historical data and current state.

Airlines use predictive maintenance to cut maintenance costs, transform non-scheduled maintenance into scheduled maintenance and to minimize undesirable ground times [24]. The potential for savings is huge. Maintenance accounts for around 10 % of airline operational costs and for almost 50 % of delays. Taking and aircraft out of service for maintenance can result in additional costs reaching 10 thousand USA per hour. Decrease in ground times via more effectively planned maintenance can bring an airline a significant financial benefit [25]. This approach has been used in engine maintenance for quite some time but airlines can perform predictive analysis more effectively today because new aircraft produce much more data and the tools for their processing are already very advanced [24]. By equipping critical components of aircraft with sensors, we are able to transmit information to engineers on ground who can get ready for a particular task beforehand and start with maintenance immediately after landing [25].

In the field of predictive maintenance we can observe the real boom of the Internet of Things because new aircraft are basically flying servers [24]. The newest aircraft types Boeing 787 Dreamliner and Airbus A350 are connected to the network even more than the passengers [25]. Boeing 787 produces around 500 GB of system data during a single flight on average while Airbus A350 is equipped with almost 6000 sensors in different parts of the aircraft creating approximately 2.5 Tb of data per day and it is expected that its new version coming in 2020 should produce 3 times more data than that [26]. The biggest airliner, Airbus A380, which first entered service 10 years ago, is equipped with 25 thousand sensors monitoring around 200 thousand parameters [25][27]. Its new version, A380-1000, designed for transport of up to 1000 passengers, will be equipped with 10 thousand sensors on each wing [26].

In the field of predictive maintenance, the so called unsupervised anomaly detection is used. It is used when we are not completely sure what we're looking for. Anomaly detection is a process of discovery which helps us determine what is happening and what we should be focused at. The program for anomaly detection has to discover interesting patterns and connections in the data itself which is done by first identification of the most important aspect – what is normal. As soon as the model achieves that, the program is able to detect outliers from the normal data set and mark them as anomalies, which are usually indications of a behavior leading to a malfunction [28].

C. Decision making support at an operations control center

The main purpose of an operations control center is an operational control of and airline and solving of operational irregularities. The main task of a dispatcher is to correctly analyse the non-standard situation, assess what new conditions and restrictions result from it and make a quick and right decision leading to sustainability of current traffic and in the best case to minimization of negative impacts including delays and additional costs.

Airline economics states that and aircraft only makes profit during flight only. Considering low margins in the aviation industry, it is an effort of every airline to maximize aircraft utilization. During a summer season, utilization of every aircraft in the fleet often reached 20 hours of net flying time a day leaving a narrow manoeuvring window for cases when one of these highly utilized aircraft becomes unserviceable due to technical malfunction for example.

In order for a dispatcher to make an optimal decision in a non-standard situation it is crucial for them to have perfect situational awareness which can only be created by timely, accurate and complete information. Among others, Big Data analysis is one of the tools which can help dispatcher to achieve that.

For sheer illustration, let's present a fictional Big Data strategy which should support decision making at operations control center in case of transferring passengers at an airport. The design of the strategy comprises following parts: *1. Big Data strategy definition, 2. Business initiatives directly supporting Big Data strategy 3. Outcomes and Factors of Success, 4. Tasks and 5. Data sources.*

The Big Data strategy will look as follows:

1. Big Data strategy definition
   a. Support and optimization of the decision making process in case of delayed flight carrying transfer passengers
2. Business initiatives directly supporting Big Data strategy
   a. Creation of an integrated information source processing all information relevant to optimal decision making and reacting to changing conditions.
   b. Increasing situational awareness of dispatchers
3. Outcomes and factors of success
   a. Reaching optimal delays in the network of the company Travel Service
   b. Increasing the probability of successfully delivering luggage of transfer passengers
   c. Improving image of the company and customer satisfaction
   d. Increasing overall success in transfer of transfer passenger up to 90 %
   e. Decreasing costs off missing connecting flights to 50 % of the current state
4. Tasks
   a. Collecting needed information from all relevant
   b. Integration and analysis of all collected information and assessing the impact of every single flight on the basis of its current status on the whole situation in the network
   c. Suggesting measures for every single flight in order to optimize delay in the network
   d. Assessing economic factors for cost minimization

e. Predicting the impact of the current state on the future development in the network

5. Data sources

Operational messages, Information about transfer passengers, Information about connecting flights, Information about ATC slots, Crew duty limitations, Information about other flights to respective destinations, Historical business data.

4 CONCLUSION

Airlines in the deregulated environment often balance on a narrow line dividing financial profit and loss. Big Data most probably represent one of the best tools for support of airlines' business strategy today which carry the potential to turn this financial balancing to the side of profit. Since the aviation industry has massive volume of historical and newly generated data at its disposal, an approach based on data appears to be tailored to it. Airlines already own the data. Now is the time to come with an innovative approach to their usage which is only limited by our own imagination and creativity while combining different data sources to find correlations.

Airlines traditionally collect big amount of transactional data which represent the basic level of Big Data analysis usage. It is already within the power of airlines to deliver personalized product based on customer profiles created from customer data which will enable offering of highly specified product packages to individual customers. A customer will get exactly the amount of service they require and which can even strengthen their brand loyalty.

The effort of every airline is to maximize their fleet utilization. This plan is, however, often disturbed by the events of non-scheduled maintenance which usually results in propagation of a delay across the airline's network and additional costs. Big Data in this case help via predictive maintenance which is based on anomalous behavior detection in data collected from sensors placed on different components of an aircraft and transferring non-scheduled maintenance into scheduled maintenance shortly before component failure. Apart from the fact that it will be possible to better predict undesired operational events, maximize fleet utilization and decrease delays cause by technical malfunctions to a minimum, we will also be able to better predict consumption of scarce sources during maintenance and manage spare parts inventory more rationally.

When it's necessary to solve operational events from the position of an operations control center, it is necessary to work with timely, accurate and complete information which are relevant to the optimal solution of the whole situation. Due to the fact that a situation can be influence by many different factors at the same time it is not always within the power of a dispatcher to assess the situation optimally. When we have enough operational data at our disposal we can instantaneously find the best solution via Big Data analysis considering the current and future state of an airline's network.

An airline is able to keep its competitive advantage when it's offering a unique product valuable to customers and difficult to be imitated by competition. It is obvious that Big Data represent a tool which can be used by many different airlines and thus isn't unique or non-imitable. It can be expected that Big Data will bring the biggest competitive advantage to companies which will be able to leverage their potential in the early phase of their development. In the future, however, they will become a common standard that will bring actionable insights and solution suggestions for future development optimization.

REFERENCES

[1] Mayer-Schonberger, V. – Cukier, K. Big Data. 1st ed. Brno: Computer Press, 2014, 256 p. ISBN 978-80-251-4119-9.

[2] The Information Age and the Printing Press: Looking Backward to See Ahead. RAND Corporation [online]. Available at: http://www.rand.org/pubs/papers/P8014/index2.html

[3] The History and Evolution of the Internet, Media, and News in 5 Infographics. ACI [online]. Available at: http://aci.info/2013/10/24/the-history-and-evolution-of-the-internet-media-and-news-in-5-infographics/

[4] Shifting the Storage Paradigm (Part One): The Evolution of Data. DDN Storage [online]. Available at: http://www.ddn.com/blog/shifting-storage-paradigm-part-one-evolution-data/

[5] The Evolution of Data. Visual Capitalis [online]. Available at: http://www.visualcapitalist.com/evolution-of-data/

[6] Han, H. – Yonggang, W. - , Chua Tat-Seng, CH. – Xuelong, L. Toward Scalable Systems for Big Data Analytics: A Technology Tutorial. Access, IEEE [online]. (2):652-687 DOI: 10.1109/ACCESS.2014.2332453. Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp

[7] Bits&Bytes: The History Of Data Storage. Backupify [online]. Available at: https://www.backupify.com/history-of-data-storage/

[8] The birth of the web. CERN [online]. Available at: http://home.cern/topics/birth-web

[9] The Evolution of Mobile Technologies: 1G to 2G to 3G to 4G LTE. Qualcomm [online]. Available at: https://www.qualcomm.com/videos/evolution-mobile-technologies-1g-2g-3g-4g-lte

[10] The Internet of Things: How the Next Evolution of the Internet Is Changing Everything. Cisco [online]. Available at:

http://www.iotsworldcongress.com/documents/4643185/3e968a44-2d12-4b73-9691-17ec508ff67b

[11] Gartner Says 4.9 Billion Connected "Things" Will Be in Use in 2015. Gartner [online]. Available at: http://www.gartner.com/newsroom/id/2905717

[12] Morgan Stanley: 75 Billion Devices Will Be Connected To The Internet Of Things By 2020. Business Insider [online]. Available at: http://www.businessinsider.com/75-billion-devices-will-be-connected-to-the-internet-by-2020-2013-10

[13] „Internet Security“. Bletchley Park. The Mansion, Bletchley Park, Sherwood Dr, Bletchley, Milton Keynes MK3 6EB, the Great Britain. 2015.

[14] 'Big Data' Is One Of The Biggest Buzzwords In Tech That No One Has Figured Out Yet. Business Insider [online]. Available at: http://www.businessinsider.com/companies-not-embracing-big-data-2014-8

[15] Novovičová, J. Probability and Statistics (Pravděpodobnost a matematická statistika). 1st ed. Prague: Czech Technical University in Prague, 1999, 154, iii s. ISBN 80-010-1980-2.

[16] MapReduce Tutorial. Hadoop [online]. Available at: https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html

[17] Big Data Takes Off: Crunching the numbers to give passengers more comfortable and cost-efficient flights. Future of Transportation [online]. Available at: http://sciamfot.com/big-data-takes-off/

[18] Davenport, T. H. At the Big Data Crossroads: turning towards a smarter travel experience [online]. Available at: http://www.bigdata.amadeus.com/assets/pdf/Amadeus_Big_Data.pdf

[19] The Evolution Of Customer Experience: Why Customer-Centric Airlines Will Lead the Market. Sabre [online]. Available at: http://sabre-2.hs-sites.com/experience

[20] Big Data Revolution: Travel Industry leveraging Big Data for Competitive Advantage. Ascend [online]. Available at: http://www.ascendforairlines.com/2013-issue-no-4/big-data-revolution

[21] Boeing Becoming More Data Driven after Realizing Big Data Benefits. Hortonworks [online]. Available at: http://hortonworks.com/big-data-insights/boeing-becoming-more-data-driven-after-realizing-big-data-benefits/

[22] Commercial Aviation and Aerospace: Big Data Analytics for Advantage, Differentiation and Dollars. IBM [online]. Available at: http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=SA

[23] Smarter Airlines: Creating Smarter Airlines with Big Data. IATA [online]. Available at: http://www.iata.org/events/Pages/airlinesintl-webinar.aspx

[24] New Predictive MRO Tools Cut Costs. Aviation Week [online]. Available at: http://aviationweek.com/awin/new-predictive-mro-tools-cut-costs

[25] Vittek, P. - Stojić, S. - Lališ, A. Proposal of Safety Data Quality Improvement Process for Aviation Organizations and Authorities. In Quality and Leading Innovation. Bratislava: University of Economics in Bratislava, 2015, p. 30-31. ISBN 978-617-589-103-2.

[26] Rozenberg, R. - Heralová, D. Quality in Air transport process of LOT Polish Airlines. MAD – Magazine of Aviation Development. 2015, vol. 3, no. 16, p. 20-24., ISSN 1805-7578.

[27] Bugaj, M. - Němec, V. - Rostáš, J. Maintenance Planning of Special Aircraft Equipment. In GEO Spatial Visions. Žilina: EDIS, 2015, p. 13-14., ISBN 978-80-554-1113-2.

[28] Dunning, T. - Friedman, B. Practical Machine Learning: a New Look at Anomaly Detection [online]. 1st ed. Beijin: O'Reilly, 2014, iv, 58 pages [cit. 2015-11-09]. ISBN 14-919-1160-3.

[29] Kraus, J. - Vittek, P., a Plos, V. Comprehensive emergency management for airport operator documentation. In: Production Management and Engineering Sciences, Proceedings of the International Conference on Engineering Science and Production Management (ESPM 2015), Tatranská Štrba, Slovak Republic, 16th-17th April 2015., s. 139-144. ISBN 978-1-138-02856-2.

AUTHORS ADDRESSES

[1] Ing. Eva Endrizalová, PhD.

Czech Technical University in Prague, Faculty of Transportation Sciences, Department of Air Transport, Horská 3, 128 03 Prague 2, Czech Republic

E-mail: endrizalova@fd.cvut.cz

[2] doc., Ing., Stanislav Szabo, PhD., MBA., LL.M

Czech Technical University in Prague, Faculty of Transportation Sciences, Department of Air Transport, Horská 3, 128 03 Prague 2, Czech Republic

E-mail: szabo@fd.cvut.cz

[3] Ing. Rostislav Pšovský

Czech Technical University in Prague, Faculty of Transportation Sciences, Department of Air Transport, Horská 3, 128 03 Prague 2, Czech Republic

E-mail: psovsky@fd.cvut.cz