A BIG DATA ENVIRONMENT FOR AIR TRAFFIC APPLICATIONS: FROM INFRASTRUCTURE TO VISUALIZATION

Paula Bezerra Rocha Garcia
Atech Negócios em Tecnologias S/A
Rua do Rócio 313, 5º andar, São Paulo/SP, Brasil
Phone: +55 11 3103-4600 ext. 1096, pgarcia@atech.com.br, Fax: +55 11 3103-4601

Giovanna Ono Koroishi
Atech Negócios em Tecnologias S/A
Rua do Rócio 313, 5º andar, São Paulo/SP, Brasil
Phone: +55 11 3103-4600 ext. 1096, gkoroishi@atech.com.br, Fax: +55 11 3103-4601

Fabio Seiti Aguchiku
Atech Negócios em Tecnologias S/A
Rua do Rócio 313, 5º andar, São Paulo/SP, Brasil
Phone: +55 11 3103-4600 ext. 1096, faguchiku@atech.com.br, Fax: +55 11 3103-4601

Claudinei Walker da Silva
Atech Negócios em Tecnologias S/A
Rua do Rócio 313, 5º andar, São Paulo/SP, Brasil
Phone: +55 11 3103-4600 ext. 1061, cwsilva@atech.com.br, Fax: +55 11 3103-4601

ABSTRACT

What challenges could be tackled in air traffic control (ATC) and management (ATM) domain when the recorded data is fully explored? Many industries have taken advantages of big data analytics in their businesses. Atech's air traffic systems collect and generate a great set of information everyday that could improve Brazilian ATC/ATM systems. This paper describes the development of a complete big data environment to store, analyze and visualize massive quantities of data. The data categories include airspace elements, flight plan and air traffic service messages, and radar detections. Case studies for the Brazilian scenario are presented in this study.

Keywords: Big data, Cluster infrastructure, Data analysis, Data visualization, ATC/ATM.

1. INTRODUCTION

In the last decade, the Information Technology market experienced important novelties in storage, transformation, analysis and data visualization. The internet went out of the offices and achieved mobility through the hands of the users in cell phones, tablets and other devices, generating a growing and overwhelming volume of data. All these facts combined led to the establishment of the so called "big data".

The Gartner Group has published in recent years a series of documents addressing the topic big data. In 2014, one of its researches pointed that the four fields large companies' CIOs will invest in are: mobile, cloud computing, social networks and big data (Howard, 2014). Synthetically, the document states that companies should invest in some of the aforementioned technologies in the coming years, otherwise they will be taken to the edge of the dispute by better positions in a competitive market.

The World Economic Forum (2012) asserts that big data will greatly impact society by creating new business possibilities and stimulating already existing business to further growth: "The amount of data in the world is exploding - large portion of this comes from the interactions over mobile devices being used by people in the developing world - people whose needs and habits have been poorly understood until now". They affirm that researchers and policymakers are beginning to realize the potential of turning these big data into useful information to provide services to benefit low-income populations. They say that "it is a call-to-action for stakeholders for concerted action to ensure that these data help the individuals and communities who create it".

The International Air Transport Association states that big data is not only the process of large amounts of data, but also big business opportunities that can offer an increasingly customized service for the clients (IATA, 2013). A cited example is the initiative "Know me" by British Airways, a system that can identify preferred customer from videos when they enter in the airport or in the boarding room. The system provides the employees information to offer special meals, discounts, etc. The system has induced a debate about privacy (O'Neill, 2012).

1.1. Technology

Big data is characterized not only by the huge volume of the data, but also by the diversity of the data. Data comes from a growing variety of sources, including logs, clickstreams, videos and social media. Structured data hegemony has been broken by unstructured data as text and human language, and by semi structured data as XML and RSS feeds (Russom, 2011). Relational databases with tables can no longer support all generated data, neither SQL queries can analyze it.

In order to handle this amount of data under a new context many tools and methodologies are being created. One important tool is Hadoop. "The Apache Hadoop software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Rather than rely on hardware to deliver high-availability, the library itself is designed to detect and handle failures at the application layer, so delivering a highly-available service on top of a cluster of computers, each of which may be prone to failures" (Apache Hadoop, 2015).

Hadoop has a distributed file system (HDFS) inspired on Google File System (Ghemawat et al., 2003). The data is organized in directories and files that are split in blocks (usually of 64 or 128 MB). The blocks are replicated across the machines to prevent failures. The parallel processing is done with MapReduce paradigm (Dean and Ghemawat, 2004).

Along with Hadoop there is an ecosystem of related projects, notably:

- Hive: a data warehouse that facilitates querying and managing large datasets residing in distributed storage. It provides a mechanism to project structure and query the data using a SQL-like language called HiveQL (Apache Hive, 2015);

- HBase: a scalable, distributed database that supports structured data storage for large tables (Apache HBase, 2015);
- Pig: a high-level data-flow language and execution framework for parallel computation (Apache Pig, 2015);
- Spark: a fast and general compute engine for large-scale data processing. It provides a simple and expressive programming model that supports a wide range of applications, including ETL, machine learning, stream processing, and graph computation (Apache Spark, 2015).

Apache Hadoop distribution is free and many companies offer other distribution with paid support. The most popular ones are Cloudera (Cloudera, 2015), Hortonworks (Hortonworks, 2015) and MapR (MapR, 2015).

While structured data has been stored in relational tables, unstructured data is not relational, demanding the NoSQL type processing. NoSQL stands for "not only SQL". At one hand we have relational tables, transactions processing, concurrency, consistency, and low latency; at the other we have non-relational, distributed data processing with high performance for high scalability.

Most of these recent databases can be classified in (MongoDB, 2015):

- Column Families: storage in columns of data together, instead of rows, optimizing queries over large datasets, e.g.: Cassandra, Hypertable, Amazon SimpleDB;
- Document Store: pair each key with a complex data structure known as a document, that can contain many different key-value pairs, or key-array pairs, or even nested documents, e.g.: MongoDB, Terrastore, CouchDB;
- Key Value: each single item in the database is stored as an attribute name (or "key"), together with its value, e.g.: DynamoDB, Redis, Voldemort;
- Graph Databases: used to store information about networks of data, such as social connections, e.g.: Neo4J, HyperGraphDB, Trinity.

In analysis domain, two programming languages stand out: R and Python. R is a highly extensible programming language and also an environment for statistics and graphics. "One of R's strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed." (R, 2015). "With origination as an open source scripting language, Python usage has grown over time. Today, it sports libraries (*numpy*, *scipy* and *matplotlib*) and functions for almost any statistical operation/model building you may want to do." (Jain, 2014). There are also complete tools and platforms available in the market for analysis. The Gartner Group published in 2015 a research about Advanced Analytics Platforms (Gutierrez, 2015) and the result pointed out tools as SAS, KNIME, RapidMiner, SAP and Revolution Analytics.

1.2. Initiatives

In the last years, many groups have been developing studies associating big data with aviation. Many of these studies address the development of environments for processing large volumes of data, while some explore specific topics in the area, such as predictive analytics for events, for example. Larsen (2013) worked on identification of key aviation data sets for operational analytics, discussing the challenges for aviation industry in the context of big data, and described the *masFlight*, a software developed to integrate different aviation databases. Ayhan *et. al* (2013) described a system that enables query processing and predictive analytics over streams of aviation data. The system uses a custom tool developed by Embry Riddle Aeronautical University (ERAU) and commercial of-the-shelf (COTS) tools.

Flight safety is another area in the aviation context that receives huge investments, precisely because it seeks efficiency continuously. Pozzi et. al (2011) described an ASTM (Automatic Safety Monitoring Tool) case, in which the huge volume of radar data was used to analyze the double track

events in a specific Flight Information Region (FIR) next to Rome, Italy. These events could compromise the flight safety.

Many industries have also taken advantages of big data analytics in their businesses. Big data can also bring developments for users and companies in the field of aeronautics, as asserted by Valeika (2015) in his article for Aviation Week. He affirms "the vast quantity of data across myriad parameters that a state-of-the-art aircraft generates could provide improvements in flight operations, reliability, maintenance and safety".

From the market's trends and the top-notch technologies mentioned in this introduction, it is clear the importance of investing in big data for air traffic applications. Atech's air traffic control and management systems collect and generate a great set of information that could promote improvements in systems' capabilities and new features from acquired insights. With the purpose of evaluating the power of this package of tools that compose the big data process, and also entering this market, a big data environment to store, analyze and visualize some of the Atech's air traffic data was developed.

1.3. Motivation

A huge quantity of data is generated by air traffic management and control processes. Despite all the technology already present in this industry, several questions still demand answers – and there are even new questions that could be formulated.

In air traffic control, each flight controller is responsible for one or two sectors of an area control center (ACC). However, there is a limited number of aircrafts that can be under a controller's responsibility. Decreasing the controller's workload is one of the concerns of today's global trend towards ATM/ATC systems' modernization. How is the sectors' occupation in an ACC during the day in Brazil? What about it during the night? And during a holiday or a specific day of week? Maybe two or more sectors could be aggregated during a specific period of day and maybe some sectors should be divided because of saturation.

Still on air traffic control, there are rules to guide flight plan filling and message exchange among controllers and aviation agencies (DECEA, 2012; Maguerroski, 2008). There is a departure message (known as DEP) that should be transmitted by the control tower to the air traffic control system immediately after an aircraft's take-off (DECEA, 2012). Unofficial sources say that, in practice, only a tiny percentage of flights send the DEP message. Is this information true? Could it be validated? What is the impact of non-compliance with this procedure?

As presented by Silva *et. al.* (2016), flight intent modeling for enhanced trajectory prediction can be supported by big data analyses through the accurate determination of specific flight related measures such as climbing and descent rates, en-route speeds, nominal and full operation flight envelope, etc.

Considering real flight trajectories, how do the aircrafts fly? Do they follow the airways' spatial limits? Do they fly differently according to the date? Maybe the aircraft model or the airline can exert an influence in the average trajectories. Are the average trajectories known?

After this brainstorming, it is clear that countless questions could be promoted. Therefore, this paper will focus on two case studies to start manipulating data:

- Air Traffic Control sector occupation,
- DEP message exchange.

2. METHODOLOGY

In our view, the development of a big data project compromises five phases: question's definition, infrastructure, data cleaning, data analysis and visualization. The challenges that will be tackled in this paper were presented in the Introduction and this section will describe the following four phases. Firstly, the hardware and software infrastructure of the installed big data cluster will be presented. Secondly, the available data will be explained by their category – airspace elements,

flight plan, air traffic service (ATS) messages and radar detection – and the data cleaning and preparation procedure will be commented. The next phase is data mining: how to process and analyze the data to gather useful information. The adopted solution for each case study, as well as the chosen tool, will be discussed. Finally, three visualization tools will be presented and one of these is Atech's 3D visualization framework.

2.1. Big data infrastructure

The big data environment was deployed to store and process the air traffic data. Hadoop was selected because of its capacity to process in batch great amount of data and possibility to couple other tools on it. The chosen distribution was Cloudera, because of its facility to install and manage, and also because we can opt for paid support in future projects that may require it.

The experimental cluster is composed by three machines with the specifications shown in Table 1.

Table 1: Cluster hardware nodes specification.

Machine	*O.S.*	*Processors*	*Memory*	*Storage*
1	Cent OS 6	Intel(R) Core(TM) i5-3570 CPU @ 3.40 GHZ	7.8 GB	917 GB
2	Cent OS 6	Intel(R) Core(TM) i5-3570 CPU @ 3.40 GHZ	3.7 GB	916 GB
3	Cent OS 6	Intel(R) Xeon(R) CPU E5560 @ 2.13 GHZ	7.8 GB	458GB

Machine 1 has the role of NameNode. In Hadoop, a NameNode is a master server that manages the file system. Machines 2 and 3 are DataNodes. In Hadoop, DataNodes store the files and process them when a MapReduce is executed.

Additionally, other two components were installed in the NameNode: Hive and Hue (Apache Hue, 2016). Hive is necessary to facilitate querying and analyzing data in Hadoop. Hue is a web user-friendly interface for analyzing data and Hadoop cluster's status.

2.2. Big data

Three relevant data categories in the air traffic context comprise airspace elements, flight plan and ATS messages, and radar detection.

Airspace elements consist of physical or virtual references to guide and to regulate air traffic. They include area control centers (ACC), terminal control areas (TMA), restricted areas, aerodromes, navaids, airways, radars, and weather zones (ICAO, 2005).

Flight plan and ATS messages guide air traffic control procedures. They are part of International Civil Aviation Organization's (ICAO) guidelines (ICAO, 2015; DECEA, 2012; ICAO, 2007). Among the flight plan types, the most popular are Filed Flight Plan (FPL) and Repetitive Flight Planning (RPL). In relation to ATS, some of the messages are Change (CHG), Cancellation (CNL), Delay (DLA), Departure (DEP), Current Flight Plan (CPL), Coordination (CDN), Acceptance (ACP) and Arrival (ARR).

Fused track data from air route surveillance radars bring information of timestamp, latitude, longitude, altitude, transponder code, etc. These data provide information of flights that have taken place.

The data categories considered in this project are summarized in Table 2. Typical data were generated for these categories, varying the data intervals from one week to one month depending on the region.

After gathering the raw data, it is necessary to adapt the different formats to structures that can be stored in Hadoop and consulted by Hive queries. It is important to preserve the nature and content of data: not to lose information or embed personal intention while manipulating data.

Data cleaning and pre-processing preparation took about 80% of the time spent in the development of this big data environment. In this project, this phase can be summarized in activities of parsing files to text or comma-separated values formats to facilitate the further data manipulation with Hive.

Table 2: Summary of data content in the Hadoop cluster.

Airspace elements	*Flight plan and ATS messages*	*Radar detection*
ACC's sectors	FPL/RPL	Timestamp
TMA	CHG	Lat/Lon
Restricted areas	CNL	Altitude
Aerodromes	DLA	Speed
Navaids	DEP	Heading
Airways	CPL	Transponder code
Radars	CDN	
Weather zones	ACP	
	ARR	

2.3. Data analysis

The analysis phase involves processing the data in order to answer a specific question, in other words, to extract useful information from the data. For this purpose, the R programming language was chosen together with the RStudio tool, a popular integrated development environment for R. This choice is based on the fact that R language usage is widespread in the data analysis community, especially for research and in the academy, and is now rapidly growing into the industry market (DataCamp, 2015). R is user-friendly, open-source, it has a huge number of libraries available and it makes data inspection, data processing, and graphic plotting easy.

The following lines explain the data analysis approach developed for each of the two case studies.

Case study I: ACC sectors' occupation

The data for this case study includes generated fused tracks and airspace elements corresponding to Brasília ACC (ACC-BS). ACC-BS is divided in sectors, including lower and upper altitude levels.

The solution's core is to decide whether a point, representing an aircraft track, is located inside a polygon describing an ACC-BS sector, while differentiating sectors' lower and upper altitudes. Then, the algorithm needs to aggregate points that belong to the same sector according to a pre-defined time interval. So it is possible to investigate the instant occupation over time (an hour, a day, a week). Additionally, filters can be applied to the results. The spatial data package called *"sp"* (Pebesma et al., 2015) and tools for splitting, applying and combining data from the package *"plyr"* (Wickham, 2015) were used in R.

Case study II: DEP message exchange

This case study considers Brasília TMA, an enclosed area containing Brasília (SBBR) and other aerodromes, and the available data are typical fused tracks and typical ATS messages.

The objective is to determine how many DEP messages were sent by flights that really took off from Brasília TMA. From fused tracks, those flights are counted by aggregating the detections in unique flights and counting the ones which: 1) first altitude in time is inside TMA's limits; 2)

flight level expectancy is ascending. The result represents the flights that have actually taken off. On the other hand, from ATS messages, the number of DEP messages with origin in one of Brasília TMA's aerodromes is also counted and the result is recorded. Finally, both results are compared calculating the percentage difference between them.

2.4. Visualization

Visualization is one of the most important phases in big data, because it has the power of transforming abstract data into visual variables. An attractive and clear way of showing data can lead users to reach extra conclusions. A tool that creates significant visualization information is more likely to be used for a longer period of time by the user, aggregating value to his/her analysis.

Atech has developed its 3D visualization framework in which it is possible to see the globe and to plot airspace elements, flight routes, flight trajectories, areas (polygons) and georeferenced images. The tool allows a complete situational awareness of the study scenario.

Some examples of airspace elements plotted in Atech's 3D visualization framework are exemplified in Figure 1.

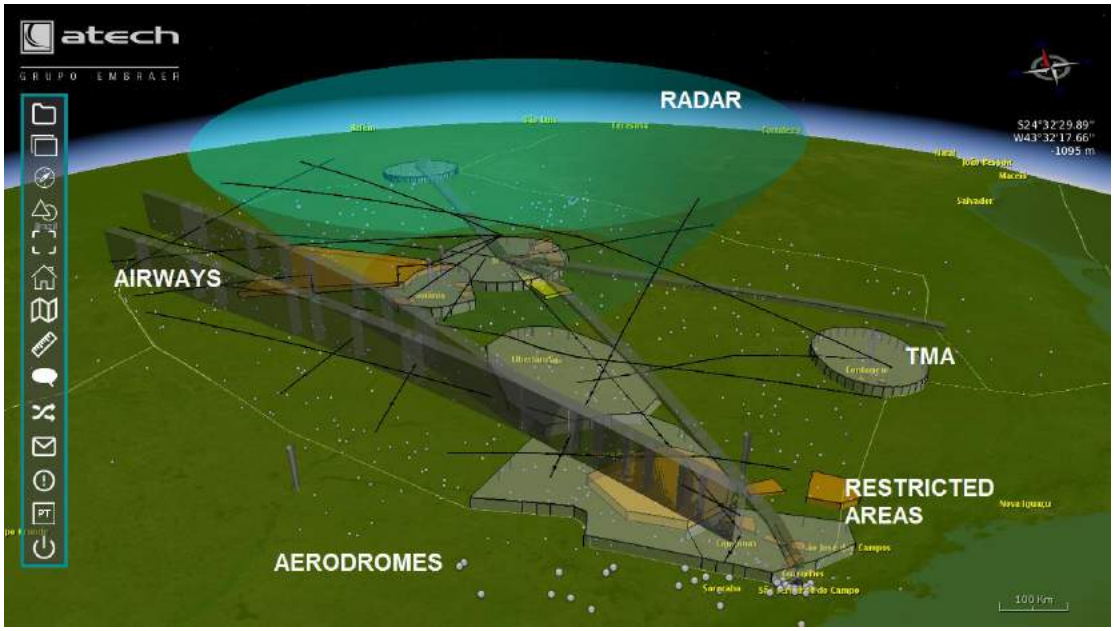


Figure 1: Examples of airspace elements plotted in Atech's 3D visualization framework: aerodromes, airways, radar, TMAs and restricted areas.

Besides the framework described previously, there are also two positively evaluated tools available in open-source community: D3 (d3.js) and R's graphics. D3 is a JavaScript library for visualizing data with HTML, SVG, and CSS (D3, 2015). It is useful for producing interactive visualizations. In relation to R, it also has several packages for robust plotting and maps integration, as *ggmap* (Kahle and Wickham, 2013), *lattice*, *ggplot2*, *rggobi* (Kabacoff, 2015).

3. RESULTS

The results of the two case studies – ACC sectors' occupation and DEP message exchange – are reported in this section.

3.1. Case study I: ACC sectors' occupation

ACC-BS is divided in 32 sectors, with 15 of them located at lower level (altitude from 0 to 149 FL) and 17 of them located at upper level (altitude from 150 to 999 FL), as shown in Figure 2.

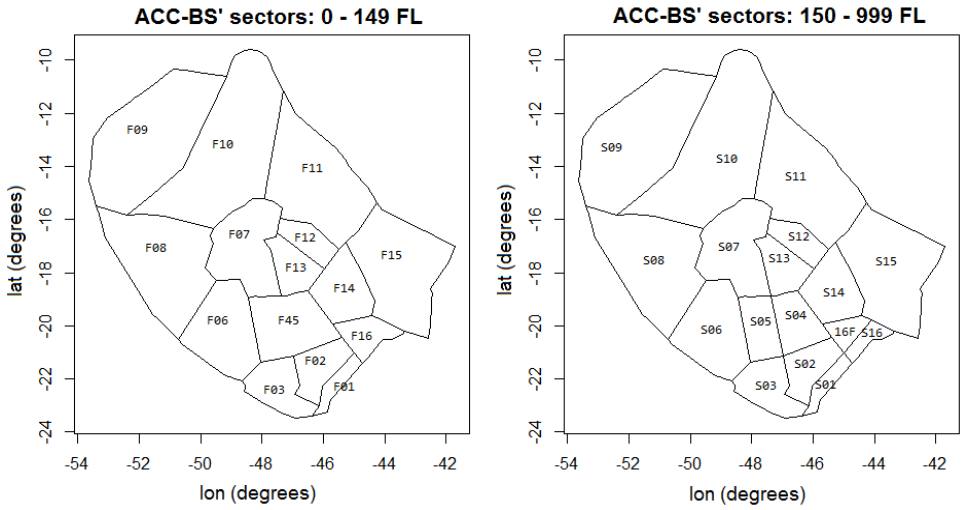


Figure 2: ACC-BS' sectors: lower and upper altitude levels.

In this case study a two-dimensional scenario is considered, so lower sectors and upper sectors are grouped, as represented in Figure 3.

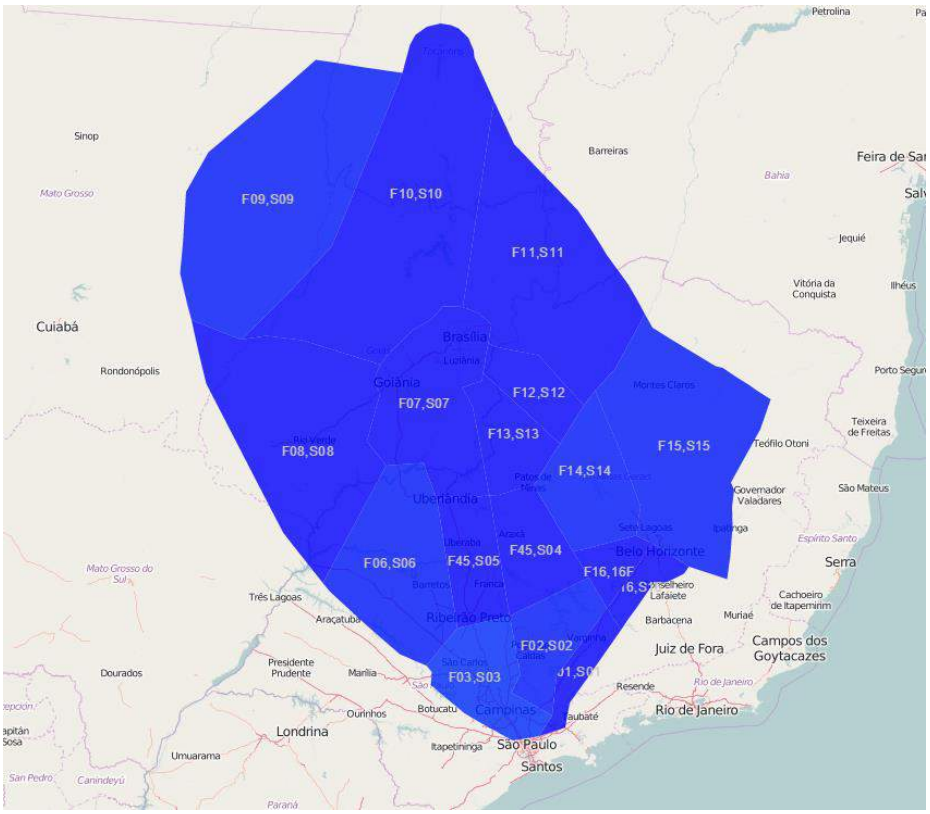
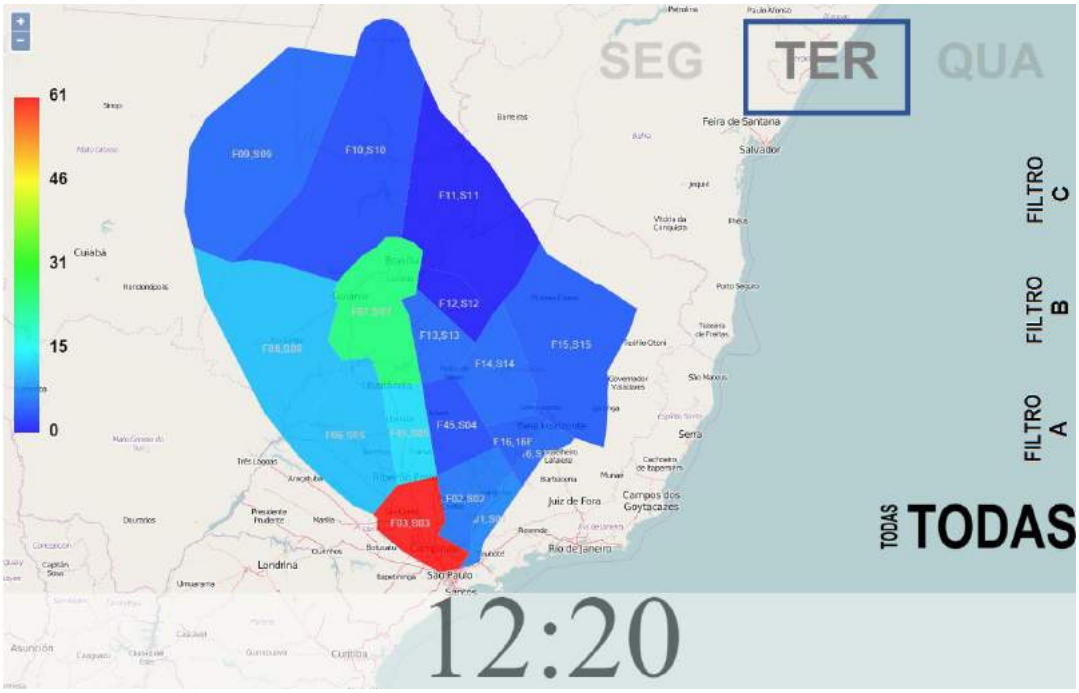


Figure 3: To represent ACC-BS in a two-dimensional scenario, lower sectors and upper sectors are grouped.

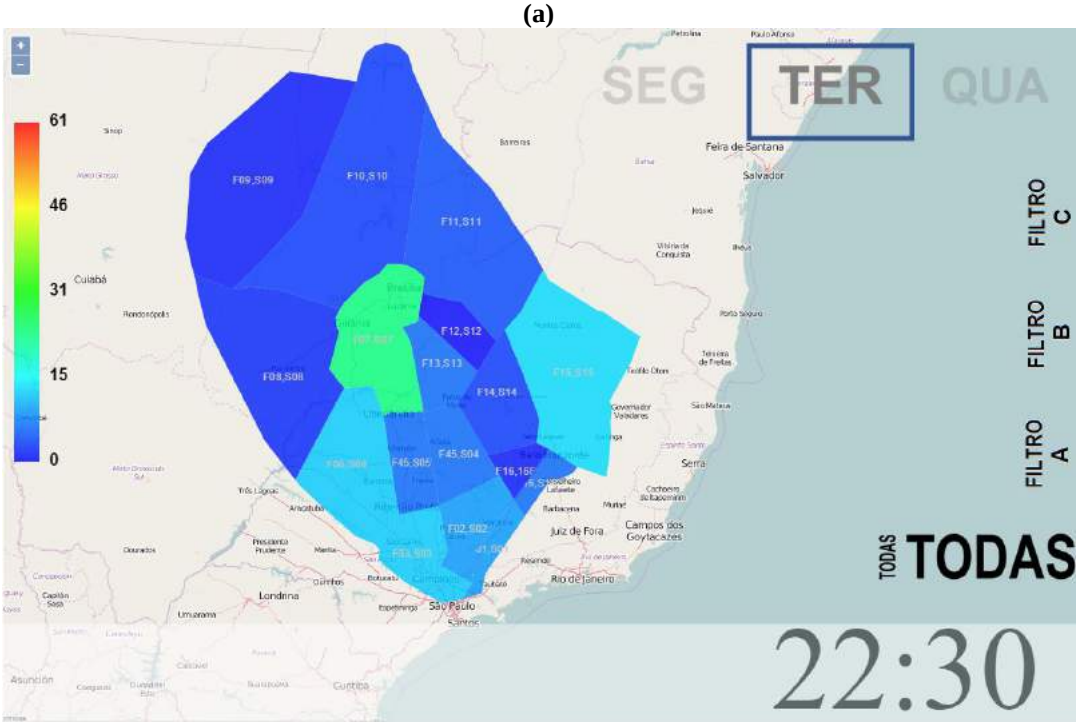
The simulated track data add up to 33 million radar detections and correspond to a simulated period of a week. The analysis phase was responsible for relating each track to a grouped sector.

In the visualization phase, an interactive user interface was developed to provide information about instant sectors' occupation. It was implemented using D3 (D3, 2015) and OpenLayers' map (OpenLayers, 2016), and it runs on a web browser. Additionally, the interface was designed to be controlled by Kinect (Microsoft, 2016), although user can interact with the application using mouse buttons. Basically, the panel contains the map, a timeline slider in UTC format, a filter to choose data period (a day of week or the whole week), other three generic aircraft filters and a color bar indicating the number of aircrafts. The language is Portuguese.

An arbitrary weekday was chosen in two different moments of the day: (a) 12:20 p.m. and (b) 10:30 p.m. Instant occupations are shown in Figure 4.



(a)



(b)

Figure 4: Instant sectors' occupation for a Tuesday at two different moments of the day: (a) 12:20 p.m. and (b) 10:30 p.m. It considers all the aircrafts.

In Figure 4a the time is equivalent to 9:20 a.m. in Brazil and one can observe an intense aircraft flow in area 3, which is near to Sao Paulo. Area 7, near to Brasília, is the second most populated region. On the other hand, in Figure 4b, time is equivalent to 7:30 p.m. in Brazil and the reduction of aircrafts in area 3 is substantial. Comparing the other areas in Figures 4a and 4b, there were an increase in the instant occupation of area 15, which is near Belo Horizonte.

Instant sectors' occupation in ACC-BS in a day of the weekend – specifically, a Saturday – is shown in Figure 5, adopting the same timetable of Figure 4a.

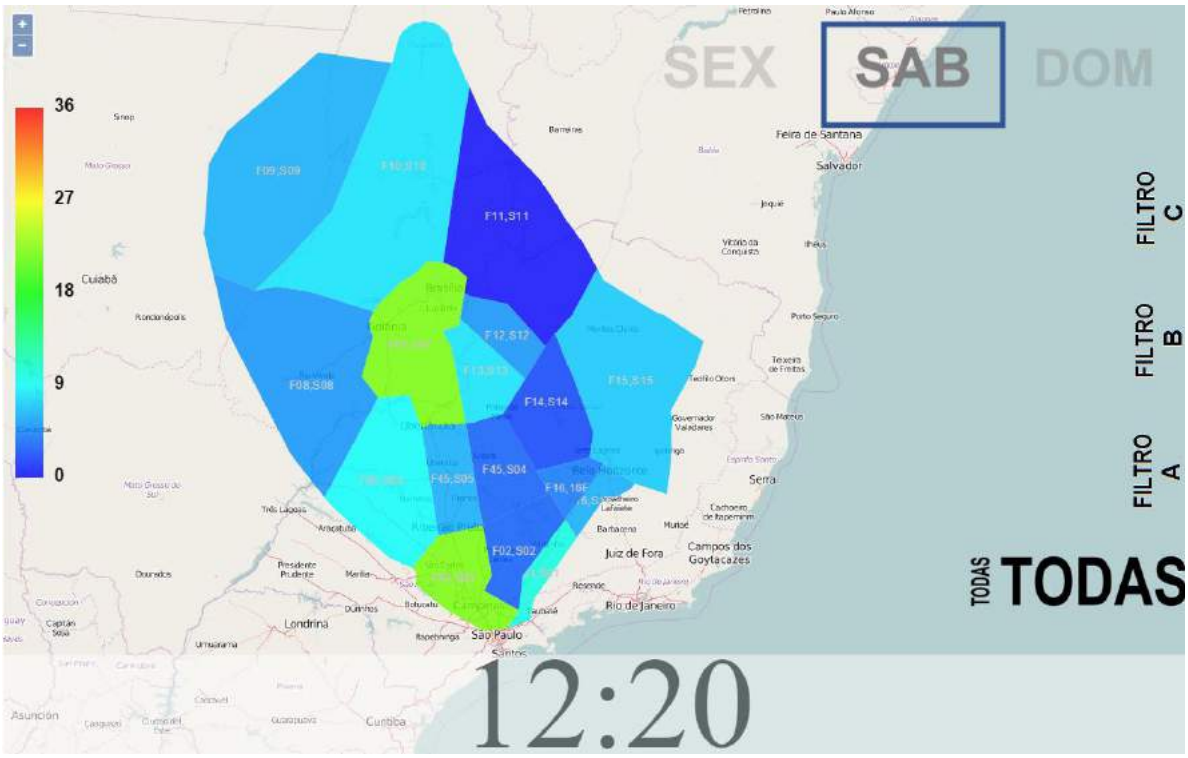
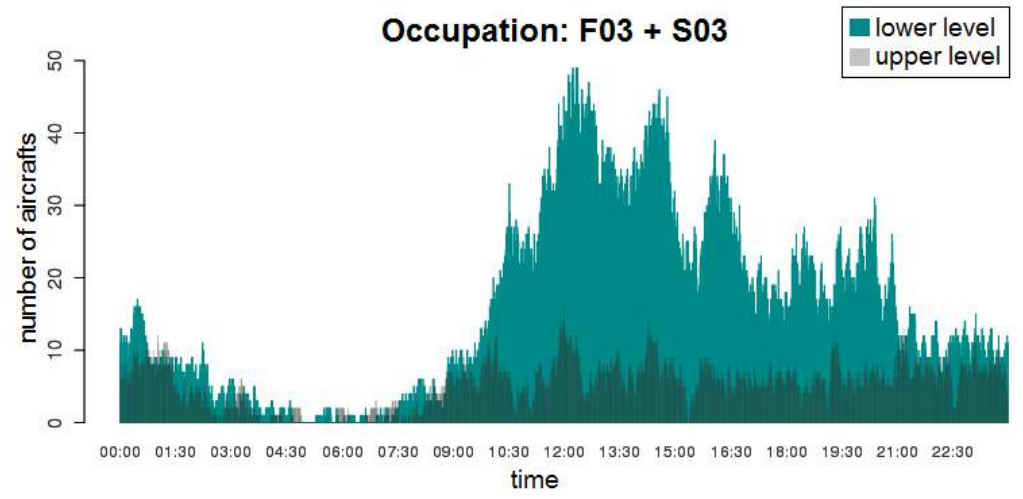


Figure 5: Instant sectors' occupation for a Saturday at 12:20 p.m. It considers all the aircrafts.

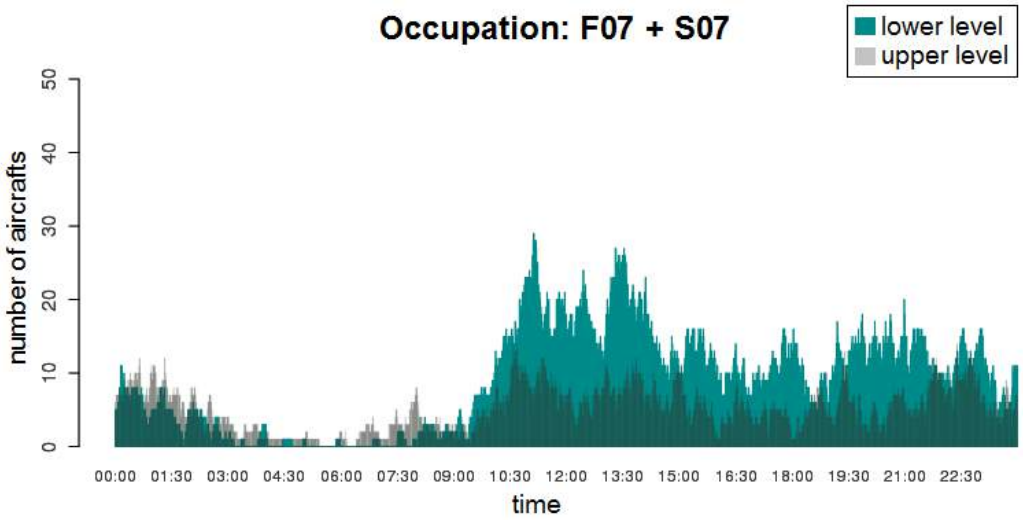
During the weekend, as shown in Figure 5, the volume of air traffic reduced considerably and it can be noticed by comparing scales of Figure 4a and Figure 5.

As the interactive interface loses information about sectors' altitudes, graphics were generated with R to split data vertically and easily check the influence of each level. Three areas were chosen in an arbitrary weekday: (a) sectors F03 and S03, near Sao Paulo; (b) sectors F07 and S07, near Brasília; and (c) sectors F10 and S10, above Brasília surroundings. The number of aircrafts over time are shown in Figure 6 for the three areas during a day.

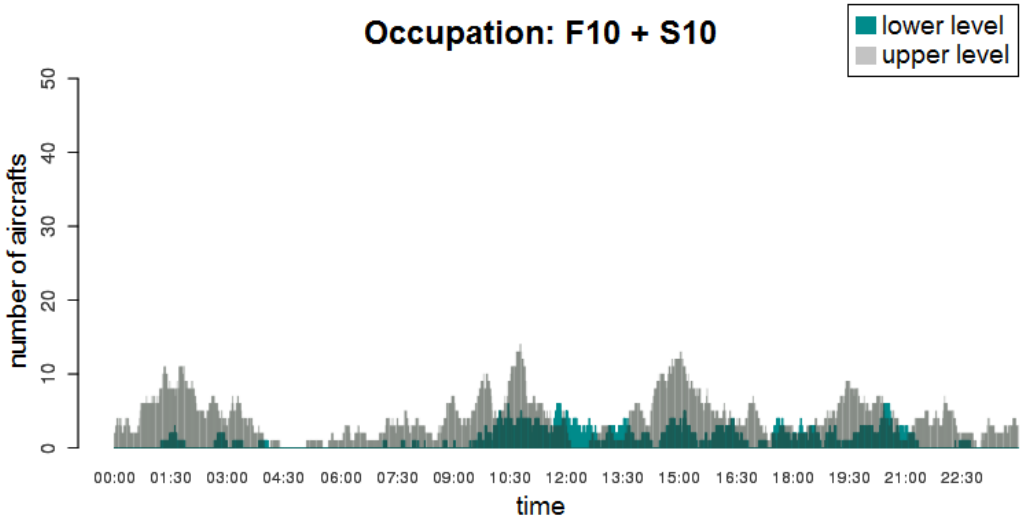
A controller is usually responsible for about 13 aircrafts simultaneously, so the results of Figures 4 and 5 could cause a misunderstanding. Figure 6, however, brings an interesting view of sectors' occupation, because it is possible to see clearly that lower sectors have a great impact on the total number of aircrafts. It means that APPs and TMAs are being considered, which includes visual aviation and aircrafts on hold. Analyzing only upper sectors' influence in the graphics, one can see that the numbers do not exceed 20 aircrafts, corroborating controller's procedures in ACC level.



(a)



(b)



(c)

Figure 6: Number of aircrafts over time during a weekday for (a) sectors F03 and S03, near Sao Paulo; (b) sectors F07 and S07, near Brasília; and (c) sectors F10 and S10, above Brasília surroundings.

3.2. Case study II: DEP message exchange

Besides SBBR, Brasília TMA comprises other 16 aerodromes, including helipads. The region is adjacent to Anápolis TMA and its shape is filled with dark gray in the route chart of Figure 7 (DECEA, 2015a). Brasília TMA has a maximum altitude of 195 FL (DECEA, 2015b).

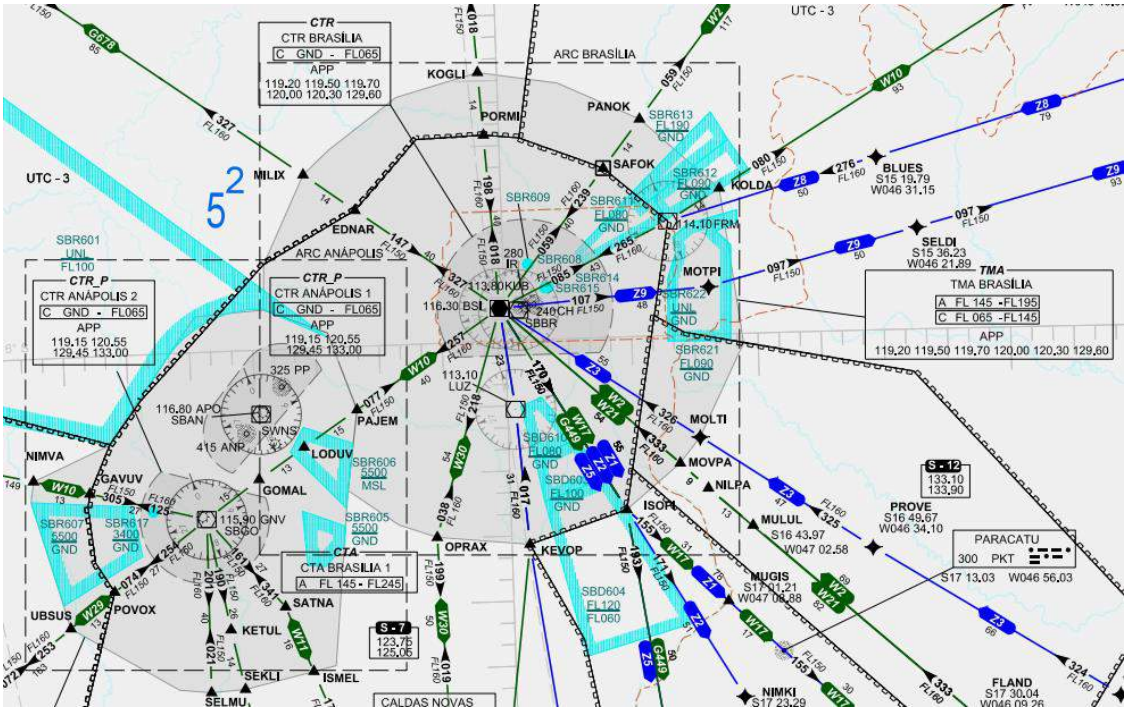


Figure 7: Aeronautical route chart ENRC L2, which contains Brasília TMA (DECEA, 2015a).

From a week of simulated radar detection data, which represent more than a million fused tracks, the detections were grouped in flights and only the flights that took off from Brasília TMA were recorded, as illustrated in Figure 8.

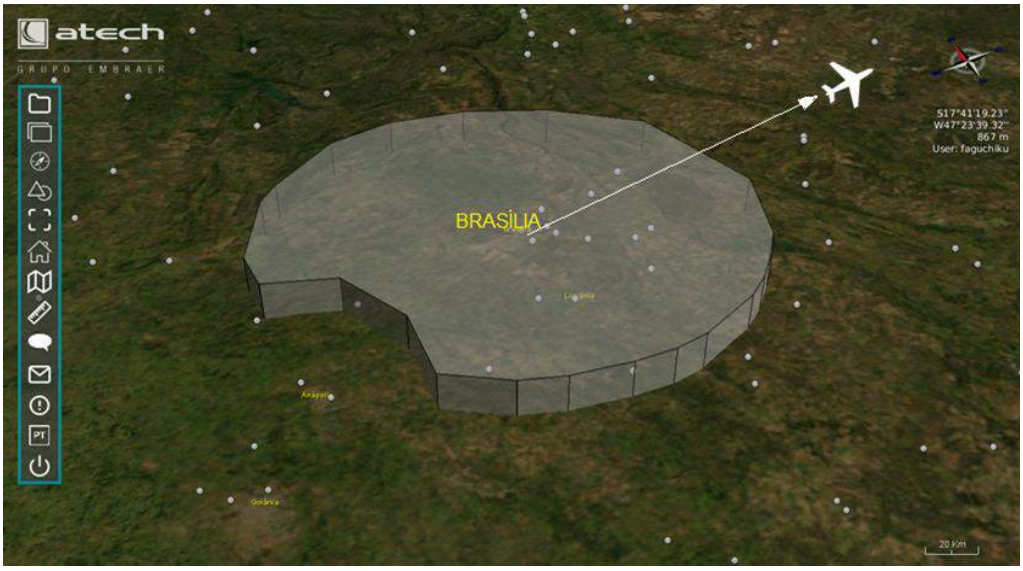


Figure 8: Brasília TMA 3D representation, highlighting Brasília aerodrome and the desired flight orientation.

The task of grouping tracks in unique flights is not that simple, because it is necessary to cross timestamp and track id. The track id is a reusable value. Moreover, flights' classification is a hard task since flights can take off and land without leaving TMA's limits, or flights can just cross TMA. The quantification of flights are presented in the Table 3.

Table 3: Flights' quantification in a week: total of tracks, total of flights and flights that took off from Brasília TMA.

Number of tracks	*Number of flights*	*Number of flights that took off from Brasília TMA*
100%	0,49%	0,19%

The simulated ATS messages' dataset was filtered using Hive queries to isolate only messages of type DEP. These messages are composed by the fields outlined in Figure 9.

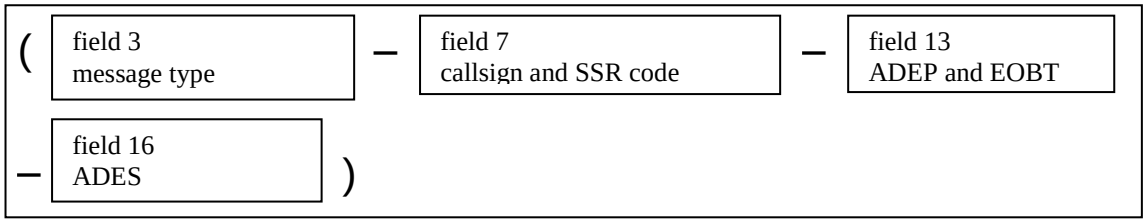


Figure 9: DEP message composition. Callsign is the flight identification; SSR code is the aircraft's transponder code; ADEP is the departure aerodrome; EOBT is the estimated off-block time; and ADES is the arrival aerodrome (DECEA, 2012).

The next step is to separate, using R, DEP messages with ADEP corresponding to any of the Brasília TMA's aerodromes, as well as to discard repeated messages. The messages include commercial, visual, general and military aviation.

After this data treatment, the resulting total number of DEP messages found and results in Table 3 were related using a relative percentual difference. The obtained percentage showed that 91.6% of flights that really took off from Brasília TMA sent the DEP message. It was expected that this number would be very low, because it is conjectured that the operational procedure requiring DEP message's submission to the responsible authority is not fully complied with. However, the result showed that Brasília TMA is really well-behaved in this simulation.

4. COMMENTS AND CONCLUSION

A big data environment to store, analyze and visualize air traffic data was developed. The data categories included in the experimental Hadoop cluster were airspace elements, flight plans, ATS messages and radar detections.

Firstly, the general perception of the technologies used is that Hadoop and Hive met our needs very well, because we were able to store our data and handle our tests using basic queries. Now it is possible to evolve our experiments by studying the use of these tools in the cloud. Secondly, in the analysis domain, R is a powerful tool especially because of the huge quantity of packages available. However, in terms of performance, R is not so fast because it is single-threaded, leading us to search for options of processing data inside Hadoop – maybe using an R-Hadoop connector or programming map and reduce routines in Java. And thirdly, about visualization, Atech will continue improving its 3D visualization framework and exploring interactive interfaces, because that is the direction to where user experience points out today (Griffin *et. al.*, 2015; Hasan and Yu, 2015; Moran *et. al.*, 2015; Ionescu 2011; Bourgois *et. al.*, 2005).

The big data initiative will persist at Atech and more data will be injected in the cluster. Typical air traffic data from other regions and from other periods of time will be generated, as well as it is planned to work with a different data context.

With respect to the case studies presented in this article, future activities are suggested for both. "ACC sectors' occupation" case could be reproduced for the other three Brazilian ACCs (Curitiba, Recife and Amazônico) and also consider a 3D map on the interactive visualization to cover lower and upper sectors. Besides it, a case study to determine average trajectories of aircrafts has already been initialized. The methodology consists of obtaining a heat map of the flights over

the territory and applying an image processing technique to obtain average values of the trajectories. On the other hand, for the "DEP message exchange" case study, it is planned to apply the method for other Brazilian TMAs – specially the less congested ones – to check if the conclusion remains. Another task is to investigate other manners of grouping tracks in unique flights and classifying flights directed outwards the desired area.

Finally, one can observe that most part of the references presented in this article are from the Internet. Discussions about big data and big data technologies are very recent. The theme involves more and more the industry and a lot of work is emerging from open source communities, justifying the huge content available on the web.

5. REFERENCES

AYHAN, S.; PESCE, J.; COMITZ, P.; SWEET, D.; BLIESNER, S.; GERBERICK, G. "Predictive analytics with aviation big data". Integrated Communications Navigation and Surveillance (ICNS) Conference, April 2013.

Apache Hadoop. Available at: <https://hadoop.apache.org/>. Accessed on 11 Dec. 2015.

Apache HBase. Available at: <http://hbase.apache.org/>. Accessed on 09 Dec. 2015.

Apache Hive. Available at: <http://hive.apache.org/>. Accessed on 09 Dec. 2015.

Apache Hue. Available at: <http://gethue.com/>. Accessed on 14 Jan. 2016.

Apache Pig. Available at: <http://pig.apache.org/>. Accessed on 09 Dec. 2015.

Apache Spark. Available at: <http://spark.apache.org/>. Accessed on 09 Dec. 2015.

BOURGOIS, M.; COOPER, M.; DUONG, V.; HJALMARSSON, J.; LANGE, M.; YNNERMAN, A. "Interactive and Immersive 3D Visualization for ATC". 6th ATM Seminar USA-Europe, 2005.

Cloudera. Available at: <http://www.cloudera.com/content/www/en-us.html/>. Accessed on 09 Dec. 2015.

D3. Available at: <http://d3js.org/>. Accessed on 10 Dec. 2015.

DataCamp. "Choosing R or Python for data analysis? An infographic". Available at: <http://blog.datacamp.com/r-or-python-for-data-analysis/>. Accessed on 10 Dec. 2015.

DEAN, J.; GHEMAWAT, S. "MapReduce: Simplified Data Processing on Large Clusters". Available at: <http://static.googleusercontent.com/media/research. google.com/pt-BR//archive/mapreduce-osdi04.pdf>. Accessed on 2015/09/25.

DECEA - Departamento de Controle do Espaço Aéreo. "ICA 100-15: Mensagens ATS". Ministério da Defesa Comando da Aeronáutica, 2012.

DECEA - Departamento de Controle do Espaço Aéreo. "ENRC L2", 12 Nov. 2015a. Available at: <http://servicos.decea.gov.br/arquivos/ais/cartas/xxxs_enrc-l2_enrc_20151112.pdf?CFID=83ed398a-d399-4cec-8404-746e11671b55&CFTOKE N=0>. Accessed on 13 Jan. 2016.

DECEA - Departamento de Controle do Espaço Aéreo. "Presidente Juscelino Kubitschek (SBBR)", 2015b. Available at: <http://www.aisweb.aer.mil.br/ ?i=aerodromos>. Accessed on 15 Dec. 2015.

GRIFFIN, W.; CATACORA, D.; SATTERFIELD, S.; BULLARD, J.; TERRILL, J. "Incorporating D3.js Information Visualization into Immersive Virtual Environments". IEEE Virtual Reality (VR), 2015.

GHEMAWAT, S.; GOBIOFF, H.; LEUNG, S-T. "The Google File System", 2003. Available at: <http://static.googleusercontent.com/media/research.google.com/pt-BR//archive/gfs-sosp2003.pdf>. Accessed on 25 Sep. 2015.

GUTIERREZ, D. "Gartner's New 2015 Magic Quadrant Reports are Out", 2015. Available at: <http://insidebigdata.com/2015/03/16/gartners-new-2015-magic-quadrant-reports/>. Accessed on 08 Dec. 2015.

HASAN, M.S.; YU, H. "Innovative Developments in HCI and Future Trends". Proceedings of the 21st International Conference on Automation & Computing, University of Strathclyde, Glasgow, UK, 11-12 September 2015.

Hortonworks. Available at: <http://hortonworks.com/>. Accessed on 09 Dec. 2015.

HOWARD, C. "The Nexus of Forces is Creating the Digital Business", 2014. Available at: <https://www.gartner.com/doc/2933218?srcId=1-3132930191>. Accessed on 11 May 2015.

IATA. "Unlocking the potential of Big Data and understanding customer 'DNA' is transforming airlines' services", 2013. Available at: <http://www.weforum.org/ reports/big-data-big-impact-new-possibilities-international-development>. Accessed on 12 May 2015.

ICAO. "Annex 2 to the Convention on International Civil Aviation: Rules of the Air". International Civil Aviation Organization, 10th edition, July 2005.

ICAO. International Civil Aviation Organization. Available at: <http://www.icao.int/>. Accessed on 10 Dec. 2015.

ICAO. "Procedures for Air Navigation Services: Air Traffic Management. Doc 4444 ATM/501". International Civil Aviation Organization, 5th edition, 2007.

IONESCU, D. "Gesture Control and the New and Intelligent Man-Machine Interface". 6th IEEE International Symposium on Applied Computational Intelligence and Informatics, 19–21 May 2011.

JAIN, K. "SAS vs. R (vs. Python) – which tool should I learn?", 2014. Available at: <http://www.analyticsvidhya.com/blog/2014/03/sas-vs-vs-python-tool-learn/>. Accessed on 09 Dec. 2015.

KAHLE, D. and WICKHAM, H. "ggmap: Spatial Visualization with ggplot2", 2013. Available at: <https://journal.r-project.org/archive/2013-1/kahle-wickham.pdf>. Accessed on 11 Jan. 2016.

KABACOFF, R. I. "Advanced graphs". Available at: <http://www.statmethods.net/advgraphs/index.html>. Accessed on 2016/01/11.

LARSEN, T. "Cross-platform aviation analytics using big-data methods". Integrated Communications Navigation and Surveillance (ICNS) Conference, April 2013.

MAGUERROSKI, G. "Plano de voo. Departamento de Treinamento IVAO-BR", Jun. 2008.

MapR. Available at: <https://www.mapr.com/>. Accessed on 09 Dec. 2015.

MICROSOFT. Xbox 360 + Kinect. Available at: <http://www.xbox.com/pt-BR/Kinect/Home-new>. Accessed on 14 Jan. 2016.

MongoDB. "No-SQL Databases Explained". Available at: <https://www.mongodb.com/nosql-explained>. Accessed on 09 Dec. 2015

MORAN, A.; GADEPALLY, V.; HUBBELL, M.; KEPNER, J. “Improving Big Data Visual Analytics with Interactive Virtual Reality”. IEEE High Performance Extreme Computing Conference, 2015.
O'NEILL, S. “British Airways and the Know Me saga: should companies run image checks on customers?”, 2012. Available at: <http://www.tnooz.com/article/british-airways-and-the-know-me-saga-should-companies-run-image-checks-on-customers>. Accessed on 13 May 2015.
OpenLayers. Available at: <http://openlayers.org/>. Accessed on 14 Jan. 2016.
PEBESMA, E.; BIVAND, R.; ROWLINGSON, B.; GOMEZ-RUBIO, V.; HIJMANS, R.; SUMNER, M.; MACQUEEN, D.; LEMON, J.; O'BRIEN, J. “sp: Classes and Methods for Spatial Data”, 2015. Available at: <http://cran.r-project.org/web/packages/sp/index.html>. Accessed on 10 Dec. 2015.
POZZI, S.; VALBONESI, C.; BEATO, V.; VOLPINI, R.; GIUSTIZIERI, F.M.; LIEUTAUD, F.; LICU, A. “Safety Monitoring in the Age of Big Data. From Description to Intervention”. Ninth USA/Europe Air Traffic Management Research and Development Seminar (ATM2011), 2011.
R. Available at: <https://www.r-project.org/>. Accessed on 09 Dec. 2015.
RUSSOM, P. “Big Data Analytics”. TDWI Best Practices Report, Fourth Quarter, 2011.
SILVA, S.; FATTORI, C. C.; OLIVEIRA, G. P.; SOUZA, E. C. “An analysis environment for improved flight trajectory prediction and conflict detection integration”. XV SITRAER, 2016 (submitted).
VALEIKA, R. “Opinion: “big data” from aircraft? Great. But for what?”. Aviation Week Network, 11 December 2015.
WICKHAM, H. “Plyr: Tools for Splitting, Applying and Combining Data”, 2015. Available at: <http://cran.r-project.org/web/packages/plyr/index.html>. Accessed on 10 Dec. 2015.
WORLD ECONOMIC FORUM. “Big Data, Big Impact: New Possibilities for International Development”, 2014. Available at: <http://www.weforum.org/reports/big-data-big-impact-new-possibilities-international-development>. Accessed on 12 May 2015.