

Development of a Data Processing Architecture for the Safety Management System (SMS)

Pedro Angélico
pedro.angelico@tecnico.ulisboa.pt

Instituto Superior Técnico, Universidade de Lisboa, Portugal

October 2020

Abstract

Aviation is expanding and its growth and prestige are profoundly affected by the risk level of operations, making it necessary to develop measures, methods and information management systems, that will feed the Safety Management System (SMS), a mandatory requirement described in ICAO Annex 19. The objective of this system is to monitor and control the safety level of operations. Technological evolution of aeronautical systems has led to the availability of a large amount of data from different sources, making manual analysis of information unviable.

The aim of the purposed work was to improve the way of accessing information, using all available data sources, to compile and process information in an organized manner, present the results clearly and identify trends using statistical treatment.

With the use of Pentaho, it was possible to process the data to create connection tables between the different databases. These tables were created in MySQL and hosted on an internal server. The relationship between the tables was established in PowerBI when they were imported, using flight id as the link between tables. The date is at the top of the data organization, enabling temporal filters in the DashBoard that was developed.

The results met the expected goals and implementation in Pentaho and PowerBI could be replicated to other data sources. Data transformations simplified other internal work, such as the creation of Individual Safety Information, giving rise to new ideas derived from the new organization of information.

Keywords: Aviation, Data Base, *DashBoard*, *Safety*, SMS

1. Introduction

The aviation sector is expanding due to the growth in air transport supported by global economic growth. The prestige of this industry is deeply affected by safety, where any accident becomes the focus of media attention.

Safety management for accident prevention is not a new topic in the 21st century. It started with the industrial revolution and has been implemented in the chemical industry [1], the oil industry [2] and the food industry [3]. In the aviation sector, safety management is not just done by airlines, but most all aviation service providers are legally obliged to have risk management systems. The aviation safety program guidelines at government and individual operator level are found in the SMS "Bible", International Civil Aviation Organization (ICAO) document 9859 [4]. As the structure is so complicated, within Europe, European Union Aviation Safety Agency (EASA) is responsible for collaborating with the European Commission in preparing

common safety rules and negotiating international aviation safety agreements.

Each member state has an authority responsible for air operations in its country. Despite EU regulation being automatically implemented in every European country, national authorities can have more restrictive legislation. Companies must create their own manuals and procedures. Portugália Airlines, like other companies, has its own procedures, such as P-DS-03 - Safety Reporting System [5], P-DS-05 - Risk Management [6], P-DS-06 - Flight Data Monitoring Procedure [7], P-DS-09 - Event Investigation [8] and P-DS-10 - Safety Performance Analysis [9].

With technology evolution, there is a huge amount of data generated in the systems which makes it impossible to manually analyze all sources of information so special tools have to be developed to do this.

This produces a huge amount of operation data and so much information makes its management very complex, giving rise to the concept of Big Data.

This work aims to improve access to this information, using all available data sources to compile and process the information in an organized way. It then intends to identify the fundamental variables, present the results clearly and identify trends using a statistical treatment.

2. Safety Performance

Regulatory requirements related to safety management and SMS implementation require aviation service providers to develop and maintain the means to check the safety performance in the organization.

Monitoring and measuring safety performance improves and maintains the levels accepted by the company (Acceptable Level of Safety - ALoS) as it provides the services.

Safety performance should be checked regarding safety policies and the desired objectives, i.e., the safety performance goals that are intended to be achieved and not exceeded, the Safety Performance Targets (SPTs).

Safety Performance Indicators (SPIs) that can provide information that ensures the company's resources are being applied correctly should be selected to check if the goals are being met.

The generic process of safety management performance and how it links to Safety Data Collection and Processing Systems, (SDCPS) and analysis is shown in figure 1. The relationship with safety promotion is shown to highlight the importance of information communication to the entire organization.

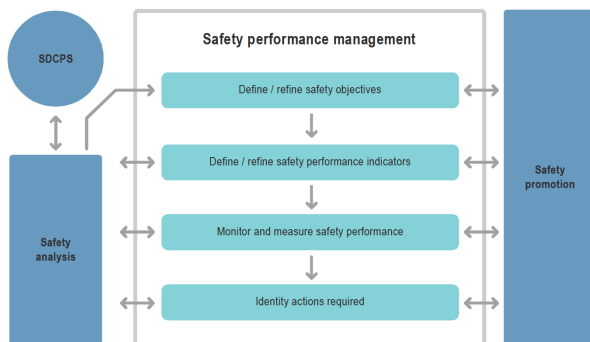


Figure 1: Safety Performance Management Process adapted from [4]

Safety performance management, through the information collected from different sources, helps the organization to answer four essential questions:

- What are the main risks in organizational safety?
- What are the objectives in terms of safety and what are the main risks that need to be addressed?

- Which direction is being followed, will it be safety objectives?
- What is the safety data and information needed to make decisions based on the information they show?

A solid database of safety information is fundamental for its management since it is the origin of data-based decision-making. Reliable safety data and information is needed to identify trends, make decisions and evaluate the safety procedure performance, relating to goals and objectives weighing the risks.

ICAO Annex 19 [10] requires that service providers develop and maintain a formal process to collect, register, act and generate feedback about the hazards in their activities, based on a combination of reactive and proactive methods to collect safety data. Similarly, chapter 8 of ICAO Annex 13 - Aircraft Accident and Incident Investigation [4] requires member states to establish and maintain an accident and incidents database for the easy and effective analysis of information about actual or potential safety debilities and to determine the necessary preventive actions.

2.1. What data to collect

Each organization needs to determine which safety data and information must be collected to support the performance management decision-making process. Data choice can be influenced by different considerations such as national and local conditions and priorities, or the need to provide data to support the SPIs monitoring.

It is important to note that the safety data and the corresponding information, which initially seem to be unrelated, may be crucial to identify problems and support data-driven decision-making.

It is advisable to optimize the amount of data and safety information, identifying what it specifically supports in terms of the effective management of the organization. Coordination between departments or divisions is necessary to optimize the efforts to create reports and data collection to avoid duplication.

Basing the information on the sources used by the company for the Portugália of Safety Risk Management [6] procedure, the various data sources are shown in figure 2.

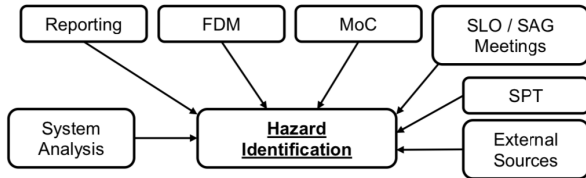


Figure 2: Identification of risk sources adapted from [6].

Due to the organization of the data and the complexity of the system, the reporting system and Flight Data Monitoring (FDM) will be focused at this point.

Reporting System

A functional and efficient reporting system is an essential part of the general monitoring function, in other words, it is a tool to identify the occasions when routine procedures have failed and the latent conditions that can lead to the recurrence of failures.

ICAO Annex 19 [10] requires member states to establish a mandatory reporting system that includes, among others, incident reports that shall be as simple as possible.

Mandatory reporting systems should aim to capture all valuable information about an occurrence, including where, when and what happened and who the report is addressed to.

Voluntary reporting systems must be established to collect data and information that is not captured through the mandatory reporting system or by the airplane systems.

FDM system

The FDM is based on the continuous analysis of data recorded during flights. This data is entered in the flight envelope to assess the proximity with the structural limits of the plane, indicated in the Aircraft Flight Manual (AFM), and with the Standard Operating Procedures (SOPs). Every time a value exceeds the limit, an alert must be triggered, the so-called Events.

FDM data of the entire company or company fleet is stored in a database system allowing the analysis of events in a general context. It can also show the evolution of the safety levels and which results are obtained with the implementation of the action, so pilot training can be adjusted to the needs.

Regardless of the data acquisition system, these must be protected according to the principles detailed in chapter 7 of ICAO DOC 9859 [4].

3. Implementation

As the implementation started, it was necessary to understand each database language and how to ac-

cess each one. Several databases are made available by selecting the one that contains information about flights, Netline as central. This database contained previous work which gave rise to a number associated to each flight with the particularity of being unique and random, Leg number (Leg_NO), which serves as a key element to connect the tables in this database. This Leg_NO was a good start for the development of this work and was the basis for linking different data. The idea of associating Leg_NO in the other databases occur, but it was necessary to find common elements that would validate a correct relationship.

Due to the amount of data, it was necessary to execute an automatic routine, trying as a first approach to creating a Python script. Once this had been created, the links to Netline were executed and some data was extracted from the database, but it was not possible to guarantee a quick and efficient solution. A Data Integration solution was then tried, which had Big Data treatment such as an Extract, Transform, Load program (ETL) that can perform several tasks in parallel [11] as the implementation basis. The Pentaho Hitachi Data Integration program was used as it is a very simple, intuitive and open source ETL tool, which allows data extraction from various types of sources with different formats, real-time visualization of the transformation and the use of Python and R script.

3.1. Netline and AQD

The flight data database, Netline, and the occurrences database, AQD, were the first link made, both with several tables. There is no direct relationship between the tables, so, it was necessary to look for common elements that allowed a correct correspondence of the occurrence to the flight. It is known that there is only one flight a day with the same identifier but for the correct match, it is necessary to compare the date.

The AQD database organization is complex but is fully organized using the occurrence number, OCC_NO, as the key.

Therefore, it was intended to create a new table with OCC_NO and Leg_NO, so that each occurrence is associated with the respective flight.

To solve the problem and simplify the search it was decided to create a key with the available data. Therefore, the occurrences database was chosen to provide only the number of the occurrence, the date of the occurrence and the flight identifier, Call_Sign. The company indicator is removed because it may have different forms such as TAP, Tap, TP, tp or even the absence of indicator. This does not add any valuable information. An example can be seen in Table 1.

OCC_NO	OCC_DATE_TIME	CALL_SIGN
O13-15	2015/07/06 09:25:00	TP1030
O16-15	2015/07/06 07:15:00	tp324
O36-15	2015/08/04 16:55:00	1468
O22-15	2015/07/23 06:50:00	Tap 1023

Table 1: Example of different type of Call Sign associated to the flight.

With the available data it is possible to create the key with the flight identifier and date, without time, with the `yyyymmddff` format. Deciphering the key through an example, 20190102412, represents the flight TAP412 of the 2nd of January 2019.

The date format is changed from `yyyy/mm/dd`, the time is removed, and the data type is changed to string using the "Select Value" block.

Letters of the flight identifier are removed and the merge from both columns is made.

On Netline, when executing the query to obtain data flight, flights are filtered to extract only data from the ones that happened, unloading a smaller amount of data. The same key was created for Netline in the same way. With the data ready to join both databases and add `Leg_NO` to the occurrences.

The first implementation for search and data association was performed using the "Join Rows" function which has shown to take about 50 minutes to run 2018 data. It was then sought a more efficient solution, the "Stream Lookup" function, which takes about 1 minute to execute the same search.

By checking the number of occurrences with the `Leg_NO` associate it was found that there were cases in which the occurrence was associated with different `Leg_NOs`. After investigating them, it was concluded that this situation was due to divergent flights. There were also occurrences without associated flights, which can be associated with maintenance, for example. A detailed analysis of the non-associated flight occurrences and events with more than one associated `Leg_NO`, revealed that these are due to flights starting one day and ending the next day.

There is a need to change the transformation and a double confirmation with the take-off and landing date, considering the take-off date as the scheduled flight departure time and landing time the real one.

These problems led to the creation of 3 branches of research: a branch for occurrences of flights that pass from one day to the next; another branch that looks in AQD flight data on Netline, and the last branch that searches the present flights in the AQD database on Netline. After the search is completed, data without a match must be filtered, then the two branches are joined at a time, starting with the last

two branches, which execute the same search but in a reverse way and lastly the branch of flights occurring the day after take-off. Before and after each merge, data should be sorted, "Sorted Merge", by the parameters to compare, before to facilitate the search, and after to eliminate repetitions, "Unique Rows". In the end, the date is added again, in the correct format `yyyy/mm/dd`, allowing it to filter by date in the future.

A new table was created with the following data: `Leg_NO`, `OCC_NO` and the date of flight departure to export the final table, "Insert /Update".

The table is stored in the Safety Department internal database, where all future tables will be kept.

3.2. Netline and AGS Events

Flight data is downloaded from the plane and loaded into an AGS analysis program, where events will be triggered. As with Netline, the data is stored in a database. In this case, it is a MySQL database type but it is only possible to access information via a SybaseIQ-type connection.

The data processing is similar to that performed in the previous point, based on the first program iteration. Data extraction begins from the source table, the unvalidated and invalid events that are not wanted, are filtered through the initial query. The same steps for creating the key are carried out to get the same results since the available data is the same. Data is filtered by no Arrival and no Departure, as it may be impossible to associate with a flight or being associated with corrupted data. Since there is no problem with the dates as in the occurrences, the search process starts and AGS provides the flight date and the event date. Again, data must be sorted before and after the "Stream Lookup", and the repeated events should be eliminated, "Unique Rows".

In the end, it can be seen that the category of events does not appear in the initial table. The event category is found in another AGS table through a "Stream Lookup", using the event descriptor for the association. The final table is exported to the safety database, "Insert/Update".

Tests have been carried out to ensure that all events were detected, by finding events without associated `Leg_NO`. After investigating the origin of these events, it was noticed that these were test or instruction flights, which were not significant to account for the company's statistical studies.

With continuous monitoring of data quality, differences began to be seen in the number of events, i.e., the number of events extracted from AGS with severity 2 and 3 was different from the number of events loaded in the new table.

Data extraction and processing were revised, and

it was realized that certain events were not considered because the comparison parameter was the date and the flight number and these did not match. The reason for this inconsistency was that the Netline database identifies flights in two different ways, Flight Number, which consists of a set of three or four numbers, without company ID, and the Call_Sign that may or may not have the identification of the company and is composed of three numbers and one letter. Unlike Netline, which always uses Call_Sign, AGS does not always use the same type of identification which required to do both distinct comparisons. The Call_Sign is the number of flight but with the last digit replaced by a letter, created when there are too many identical flight numbers as a way to prevent mistakes. Therefore, it was necessary to create a key with Call_Sign and another with Flight_NO, making the research in two branches consecutively. In both branches, events without associated flight are eliminated, they are ordered by Leg_NO and both tables are joined. After this change, a new extraction from the final table was made and, when compared to the original table, it was concluded that they were similar.

The parameters required from flight data can be chosen in this final table, just requiring the date and a flight indicator to associate with the Leg_NO.

3.3. Netline and AGS Flights

With the events and associated Leg_NOs, more data was needed to perform FDM statistical analysis. Besides the importance of studying the events, one must realize the quality of flight data, that is, how much flight data that can be extracted from the plane and how much data is lost, as well to know this information specifically by plane and by fleet.

The loss of flight data is harmful, events requiring inspection or maintenance may be lost and in the event of an occurrence, the data flight information may not be available. Flight data continuous quality monitoring should, therefore, be done to inform maintenance as soon as a failure is detected.

The previous connection cannot be used to collect this data as flights do not always have events of 2 or 3 severity, and some flights will not be covered. However, beyond the table of the events, there is a table with all the flights analyzed by the AGS. The previous program was replicated to extract this data, only changing the origin of the data, which is now the AGS event table rather than the flight table. The final table is kept in the database and is composed of Leg_NO, departure and destination airports, the flight identifier, departure date and aircraft registration number.

3.4. Crew and Flight

In the events and occurrences analysis, issues related to what happened or why a given decision sometimes arise and at these moments the best action is to directly contact the captain and the first officer. Another useful report is their experience with the aircraft in question. This information is available in the company in the Netline database tables. It was decided to associate all AGS flights with details taken from Netline that are pertinent to the analysis, departure time and arrival, to bring this information together into one table. For this first step, the table in the safety database previously created in the interconnection of Netline with AGS was used.

In another separate branch, the information of crew per flight is extracted from Netline. The Netline table is structured as follows: Leg_NO, crew member name and role. To organize information with the same structure as the other tables it was decided to use "row denormalise", thus obtaining one line per flight. There may be, on the same flight, more than one commander or first officer, when lines were checked. In these cases, to have separate license dates, the "split fields" tool was used, defining the comma as a delimiter. This step is performed for captains and first officer. Flight data and crew data come together via a "Stream Lookup". After joining, the search is made in four different phases, filtering in the query by the occupied position in the crew to speed up the search. This is executed through the name of the crew member and the type of aircraft because there are pilots with more than one license.

In the end, before exporting the table crew, repeated data is eliminated: the type of aircraft, the starting column with the two captains and first officers together. As a rule, a "Select Values" block has been left at the end, since through this it is possible to change the name and order of the columns, the data format among other factors, making it a useful tool in the future.

The table is exported, in the same way as the previous ones, containing the names of captains, first officers and the validity of each one's license.

3.5. Netline and AGS all events

The events that must be analyzed in detail are undoubtedly the most severe, but they should all be analyzed with some attention as interesting trends can be traced for training, for example. Therefore, it was decided to include a table with all events regardless of their severity.

For this implementation, the program developed for events, in general, was used changing the filter in the query, looking for validated and non-evaluated

events, excluding only the invalid.

Finally, the table is exported with the same format as the table generated in "Netline and AGS events".

3.6. Merge of all databases

The final goal of this table was the association of all the information about the events, the crew, occurrences and METARs, not only to assist future data treatment and statistical analysis but mainly to aggregate and organize all the information for the analysis of an event or specific occurrence.

The events, the crew and the occurrences are extracted directly from previously created tables and stored in the Safety Department database and information for METARs: day, time and airport are extracted from a table created by the Engineering and Operation Department. Access to this information has been requested and provided to the Safety Department to answer queries. A manually created table, airports, stored in the Safety Department database contains the airport code dictionary able to transform the ICAO code into the IATA code of airport in question.

To get METAR itself, a new research block, Database Value Lookup was used. This block consists of the same Lookup Value but extracts information as it is in the table and does not allow the data transformation at the time of the query. In this case, it is preferred to use this method if no change is necessary, extracting only the necessary data. Access to the Engineering and Operation Department database again and data is extracted to the table, with the respective departure and arrival airport METAR.

Before exporting the final table, data is prepared by removing the word "Class" and the space between the word and the number of severity through the "String Cut". Columns are organized in the intended way, "Select Values", and ordered, "Sort Rows", before removing the duplicated lines, "Unique Rows". As a last step before the exportation, lines are filtered by severity 2 and 3, excluding the remaining events.

4. DashBoard

When starting the development, there were doubts on how to create the DashBoard, through an application, from a web page or if there was a program for this. A search showed a considerable number of programs were found but the two most interesting ones were PowerBI and Tableau.

PowerBI is a Microsoft program that gets data from different sources, then transforms, views and shares it [12]. At first glance, it looks very similar to other Microsoft programs, featuring a user-friendly interface.

However, this was not the program used in the company for this purpose, Tableau [13] was used.

A program with the same functionality, with very similar tools but it was more complex and less user-friendly than PowerBI.

Both programs have a desktop, Web and mobile version so this was not a decisive factor. PowerBI was chosen, not only for its Microsoft familiar interface but also by company preference for Microsoft programs.

Another reason that led to the choice of this program was the introduction of Microsoft Azure in the company, a cloud that can work with a server and that in the future it could become the place where the databases of will be stored.

4.1. Data Input

The data input process begins with connections performed in a similar way to Pentaho. This is when one realizes the possible sources of information links that PowerBI can execute, for example, data from excel, text, PDF, among others. It also can access data in the various formats of Azure, websites, online services, Python and R scripts or even create a blank query.

After understanding how the links are made, an attempt to access the Safety Department database was made where the created tables are located. MySQL database was chosen in the link, entry credentials were inserted, and the available tables were immediately accessible.

4.2. Data Preparation

The introduction of a temporal filter raised the first question since it did not work with all tables equally, as they have the date in different formats and there can be more than one date field within the same table. Two solutions were found: change the format of the date or make a different filter.

A free filter was created, with the possibility of selecting the year, the quarter, the month or even the day. When searching for solutions to this issue, a tool was found for PowerBI, Data Analysis Expressions, DAX, which aims to solve data analysis and basic calculation. The DAX tool provides a collection of functions, operators and constants that can be used in formulas, expressions or to calculate values, i.e., it helps in creating new information from data.

Through the following DAX function (start_date end_date) it was possible to create the calendar but it was still not possible to make the selection as intended. In this function, the beginning and the end of the calendar was defined, and 2018 was chosen as the beginning for being the year of full integration of FDM software and reporting and the year

2021 as the end, to shorten the list of dates without information.

Extra columns had to be created with year, quarter, month and days of the week to add the necessary and accurate information. The code used is available in [14]. With the date running and with the ability to create visualizations, it was decided to stop the implementation and define what data would be presented and what the relationships and transformations needed to be created.

Discrimination of the number of flights performed was found relevant to understand the size of the operation and contextualize the indicators and metrics to be presented. A new link needed to be created, directly to Netline. The introduction of Netline in the data set made it possible to create a different structure.



Figure 3: Database scheme without interconnection.

Looking at figure 3, it makes sense to connect the function dates to the flight date on Netline and then make the remaining connections through Leg.NO. This figure only contains the table concerning the AQD that performs the connection between it and Netline. It was intended to create the possibility of studying the occurrences by Root Cause, Event Type, Event Descriptors, Immediate Effect, Operational Effect and ATA Code. This meant entering the AQD tables, linked to this relationship table. When entering the original tables, and so that the system of relations does not become too complex, the original table was changed and it was possible to create a new column in this table through Merge Queries with the respective Leg_NO associated with each occurrence. It was then possible to introduce the remaining AQD database tables, executing the

link to the main table.

After all the data had been entered in PowerBI, it was found that the use of filters through graphics interactivity was not working due to table formats. The information that belongs to the same occurrence, has the same OCC_NO but each piece of information is on a different line. There had to be one line for each occurrence and several columns with the desired information. Information was grouped by occurrence and the Pivoted Column function was used to obtain the intended table, one row per occurrence.

At this point, it was possible to start creating links between the tables. The calendar is linked to the Netline table associating Date with DAY_OF_ORIGIN, respectively. The Netline table is linked to the main AQD tables associating Flight Cycles to LEG_NO. The main AQD table is linked to the remaining tables in this database with relevant information. This link is executed through OCC_NO. AGS databases are both linked to Netline table via the Leg.NO. A relationship is created between the table with all events and tables with Root Causes and Event Descriptors for events with associated occurrences. The table with analyzed flight data is connected to Netline, and this is connected to the table of events.

When creating links, it is necessary to realize what is the data flow and what kind of relationships exist between data: many to one (*: 1), one to one (1: 1), one to many (1: *) and many to many (*:*) . Case by case analysis was performed to execute the best choice of relationships.

All the connections between tables were then created as can be seen in 4.

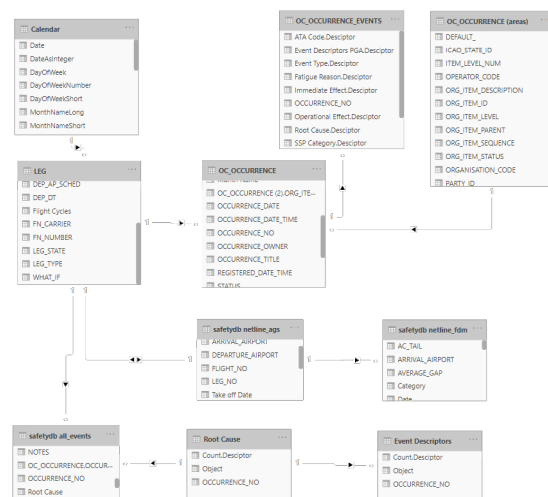


Figure 4: Schematic representation of the links between Netline, AQD and FDM.

4.3. Statistical Data

Once all the data had been connected, the graphical part of the dashboard was initiated. It was defined to include a date filter and a numeric indicator with the number of flights within the selected period that was transversal to all pages. As in the number of flights, in pages of events was also placed a numeric indicator with the number of occurrences in the selected period.

It was defined that the data concerning the occurrences are the points previously mentioned: Root Cause, Event Type, Event Descriptors, Immediate Effect, Operational Effect and ATA Code. It was also intended to carry out an individual study by airport and construct a table with data related to reporting and notifications dates, and reporting and notification times. Moving on to the FDM, the quarterly presentations were used as the basis for data presentation. This data had been presented, but until the end of this work, the data was transformed into Excel and then shown to PowerPoint for each presentation without any automatism.

An overview of the latest FDM presentations was made to build a Dashboard with the correct information and it was annotated to present the data clearly and objectively.

The purpose of the quarterly FDM meetings is to present an overview of the quarter in terms of the number of analyzed flights, the exceedances analysis, a detailed study on critical operations airports such as London or Florence, where events and consequences are discussed, and mitigation measures are analyzed. Events that are usually analyzed in detail have a direct relation to the limitations of the aircraft. Hard Landings, Maximum Operating Limit Speed (VMO/MMO), Flap Overspeed and the Go maneuver Around are analyzed.

4.4. Data display

The Data Analysis Expressions (DAX) tool once again was used to perform some analysis and statistical studies such as calculating the number of hours of reporting after events and time between reporting and authority notification, percentage and permillage calculations, to check if the occurrence associated with a Go Around flight is effectively or not related to this maneuver, checking the Go Around report, and for the study of unstated approaches.

It is presented in 1 the condition created to verify inside the filtered table by Go Arounds, if the word "go around", "go-around" or "ga" is in the occurrence title.

$$\begin{aligned}
 &IF(\\
 &\quad containsstring('data_base'[table.field], "goaround") || \\
 &\quad containsstring('data_base'[table.field], "go - around") || \\
 &\quad containsstring("data_base'[table.field], "ga"), \\
 &TRUE, \\
 &FALSE)
 \end{aligned} \tag{1}$$

Formula 1 was adapted to fit this document: data_base means safetydb all_events, table means OC_OCCURRENCE and field means OCCURRENCE_TITLE.

In the wake of the Covid-19 pandemic and the reduction of flight operations, IATA carried out a general study about the events and detected a significant increase in the number of Unstable Approaches. These pose risks to the operation and are a precursor of high-risk events.

These consist of approaches to airports that are detected by non-compliance with the company's SOPs [15] becoming important to study its reason, Root Cause.

Accessing the all_events table, a new table was created using a built-in filter, formula 2. An Unstable Approach must contain at least one of the following events: High-speed approach 1000 ft, Late Flap Setting, Late LG Extension, Localizer deviation at 1000 ft or Above or Below Glideslope at 1000 ft.

At the end of 2, the sum of zero aims to put the zero value on flights without Unstable Approach.

$$\begin{aligned}
 &UnstableApproach = \\
 &\quad calculate(\\
 &\quad\quad counte('safetydb\all_events'[LEG_NO]), \\
 &\quad\quad 'safetydb\all_events'[EVENT_DESCRIPTION] IN { \\
 &\quad\quad\quad "Highspeedapproach1000ft", \\
 &\quad\quad\quad "LateFlapSetting", \\
 &\quad\quad\quad "LateLGExtension", \\
 &\quad\quad\quad "Localizerdeviationat1000ft", \\
 &\quad\quad\quad "AboveGlideslopeat1000ft", \\
 &\quad\quad\quad "BelowGlideslopeat1000ft", \\
 &\quad\quad\quad "UnstableApproachat1000ft" \\
 &\quad\quad } \\
 &\quad) + 0
 \end{aligned} \tag{2}$$

5. Results

Regarding the occurrence's analysis, it became possible to check the events that were not yet closed, the average reporting time and investigation time for each occurrence as well as the mean, which is

important due to the legal requirement of 72 hours for authority notification.

In Figure 5 gives an example of a DashBoard page of reported occurrences. This page contains two filters, making it is possible to define the time frame to analyze and the operation area to study. The results are based on the main SPIs chosen by the Safety Department. It is not possible to demonstrate in this document, but interesting interactivity was found in the data, enabling a greater perception of associations between indicators.

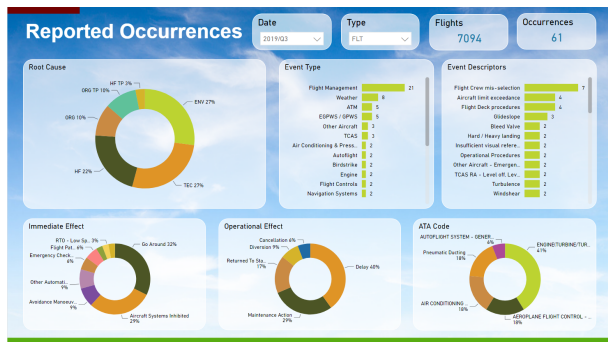


Figure 5: Page of Occurrence by airport in Dashboard created in PowerBI.

Concerning FDM data, Individual Safety Information is being developed based on the newly created tables. It was possible to start developing a web page to present the individual contribution of each pilot towards the company's statistics in real-time, as well as the events associated with each one using the transformations and relations that were created.

Quarterly FDM presentations are now almost automatic, avoiding repetitive work, increasing the time available for individual investigation of events that require increased attention. Figure 6 represents one of the Dashboard pages, with graphics of event categories, with severity and percentage, and also the temporal evolution of events with severity 3.

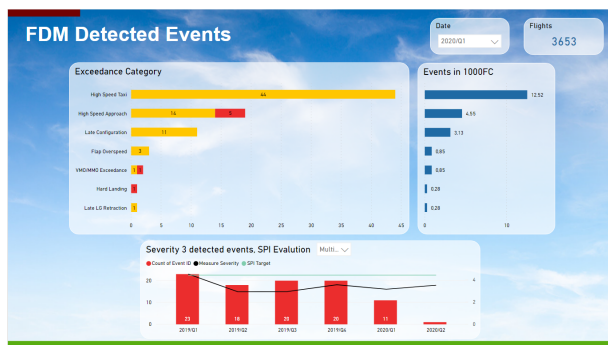


Figure 6: Page of FDM Detected Events by Airport from Dashboard created in PowerBI.

6. Conclusions

The proposed objectives were met and a data transformation was developed using Pentaho, which can be replicated for new data sources, which is an asset for the company. The tables that are created are then stored on the Portugália Safety Department server and allow numerous applications, such as the research of events or occurrences, which is the basis for the creation of Individual Safety Information, and in a next step, this information can be used by an application with individual data. Future implementation through the introduction of crew fatigue data to associate with events and occurrences.

A Dashboard was created that allows constant monitoring of operational data, as well as its evolution. The goal of a clearer, objective presentation was achieved and a new tool for the FDM meetings was generated.

This work had a positive impact on the daily functions performed at the Safety Department.

The enormous potential in works and some future applications can be proposed since both programs have an enormous capacity to expand the implemented solutions.

Acknowledgements

Thank you to everyone who made it possible to accomplish this work, the Instituto Superior Técnico lecturers, my work colleagues and all my family and friends, who supported and helped me.

References

- [1] M.N.Vinodkumar. A study on the impact of management system certification on safety management. *Safety Science*, 49(3):498–507, 2011. doi:10.1016/j.ssci.2010.11.009.
- [2] Engobo Emeseh Ambisisi Ambituunia, Jaime Amezagaa. Analysis of safety and environmental regulations for downstream petroleum industry operations in nigeria: Problems and prospects. *Environmental Development*, 9:43–60, 2014. doi:10.1016/j.envdev.2013.12.002.
- [3] Sanford A. Miller. Food safety — an international concern internationalization of food science and safety. *Food Microbiology*, 7(1):81–88, 1990. doi:10.1016/0740-0020(90)90011-6.
- [4] ICAO. *Doc 9859 - Safety Management Manual*. 4^a edition, Montreal, Canadá, 2018. ISBN:978-92-9258-552-5.
- [5] Safety Department. *P-DS-03 - Safety Reporting System*. 8 edition, 2019.
- [6] Safety Department. *P-DS-05 - Risk Management*. 2 edition, 2019. pag. 6-8.

- [7] Safety Department. *P-DS-06 - Flight Data Monitoring Procedure*. 2 edition, 2019.
- [8] Safety Department. *P-DS-09 - Investigação de Ocorrências*. 1 edition, 2019.
- [9] Safety Department. *P-DS-10 - Safety Performance Analysis*. 3 edition, 2019.
- [10] ICAO. *SAFETY MANAGEMENT: icao annex 19*. 2^a edition, Montreal, Canadá, 2016. ISBN:978-92-9249-965-5.
- [11] Olli Sjöblom. Data mining in promoting aviation safety management. *Safe and Secure Cities*, 7(1):p. 186–193, 2014. doi:10.1007/978-3-319-10211-5_19 (acedido a 19 de Outubro de 2019).
- [12] O que é o power bi desktop?, (accessed on august 22, 2020).
- [13] Get tableau desktop as part of tableau creator, ((accessed on august 22, 2020).
- [14] How to create a date table in power bi in 2 simple steps, (accessed on march 4, 2020).
- [15] Flight Operations. *Safety Management Manual*. 4 edition, 2018. (acedido em 27 de Outubro de 2019).