ICAO - AVIATION BIG DATA PROCESS

Contents

1. Introduction

The COVID-19 pandemic has profoundly impacted the operations of air carriers, airports, and air navigation service providers (ANSPs). The decisions taken today, and in the months to come, will impact the future of the industry. It is imperative for States, industries, and all stakeholders to have reliable information and tool to monitor and assess the evolving impact of COVID-19 and leverage key indicators to make informed, data-driven decisions.

ICAO, using ADS-B Flight aware data and the ICAO Enterprise Data Management (EDM), has worked to develop interactive dashboards to monitor different aspects of the impact of COVID-19 on civil aviation. This document provides a description of functionality of all COVID-19 Pandemic Traffic Data Analysis dashboards suite.

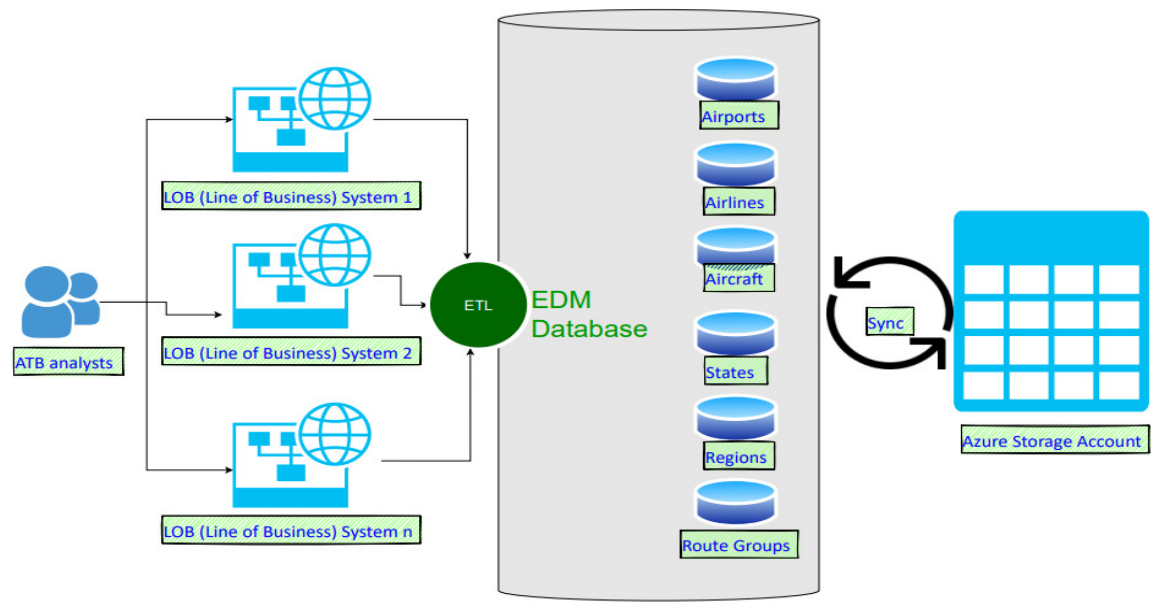# 2. Aviation big data processing model

2.1 Big data processing environment

The big data process is hosted on Microsoft Cloud Azure environment. Azure Cloud HDInsight cluster 4.0 has been used to process the big data. Raw and processed data are stored in an Azure Cloud Storage account with unlimited space. HDInsight cluster HIVE language has been used to process the raw data and generate the monthly summary files. The worker nodes within the HDInsight cluster are increased to the maximum (6) for optimal performance when processing the big data

2.2 Reference data

ICAO creates and maintains reference data using multiple LOB (Line of Business) system and EDM (Enterprise Data Management) database. Reference data includes information about Airports, Airlines, Aircrafts, States, Regions and Route Groups. Reference data is synchronized into the Azure Cloud Storage Account then used in the big data processing. ICAO reference data architect is described in Figure1.
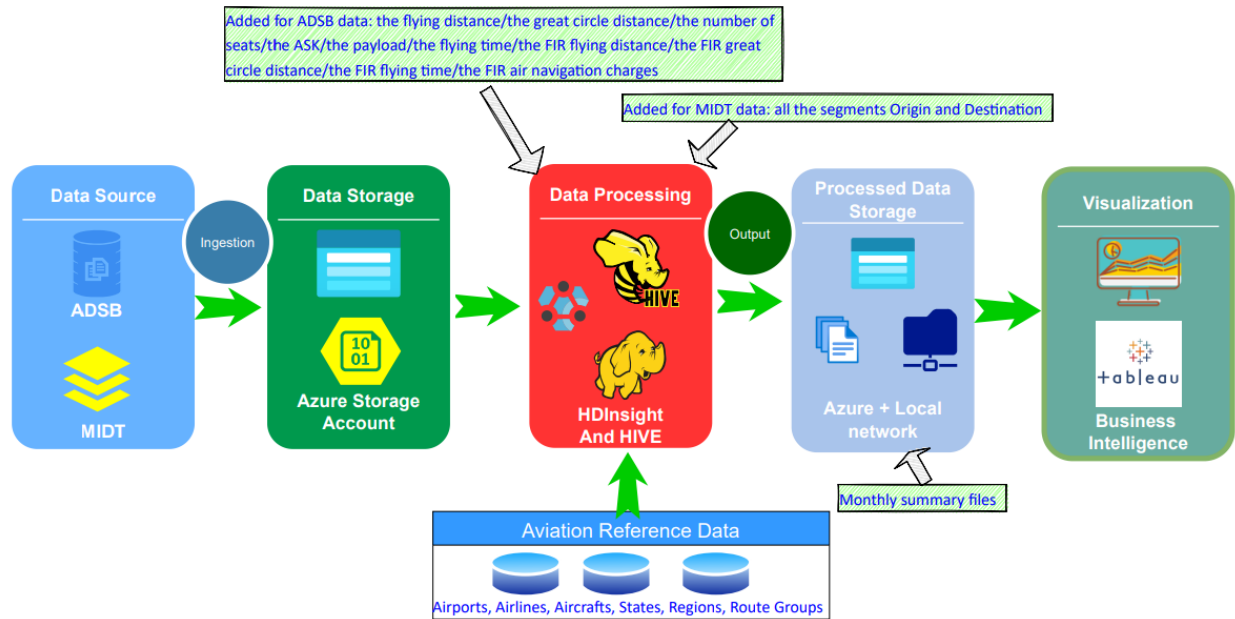
Figure 1 Reference data architecture

2.3 Big data process model

ICAO aviation big data process model includes multiple phases from data source, storage, processing, outputting, and visualization. The architect of this process is described in Figure 2.

Figure 2 Big data process model

2.4 ADS-B data process

ADS-B raw data is downloaded from FlightAware to the Azure Cloud Storage Account. ADS-B raw data include information related to each flight: Origin, Destination, Airline, Aircraft, Departure time, Arrival time. ADS-B raw data also include positions (Latitude, Longitude) for each aircraft during the flight. We have about 6 positions for each aircraft every minute. The size of the ADS-B raw data is about 3-4 million rows per flight information and more than 600 million rows of aircraft positions per month. Combined with the Reference data, HIVE language is used in the Azure Cloud cluster to calculate for each flight, the flying distance, the great circle distance, the number of seats, the ASK, the payload, the flying time, the FIR flying distance, the FIR great circle distance, the FIR flying time, the FIR air navigation charges. Combined with the Reference data, the ADS-B monthly summary files are generated in the Azure Cloud Storage Account. ADS-B Monthly summary files include information about Origin, Destination, Airline, Region, Route Group, number of flights, number of seats, flying distance, flying hours, flying ASK. ADS-B Monthly summary files are downloaded to ICAO network drive. ADS-B Monthly summary files are loaded into the Business Intelligence platform (Tableau) which analyzes the data, creates the visualizations, and distributes it to the proper audience. The metadata of the ADS-B data and the architecture of joining with the reference tables in EDM is available at [LINK].

2.5 MIDT data process

MIDT raw data is downloaded from OAG to the Azure Cloud Storage Account. MIDT raw data include information related to Origin, Destination, Gateway (2 maximum), Airline and passengers. The size of
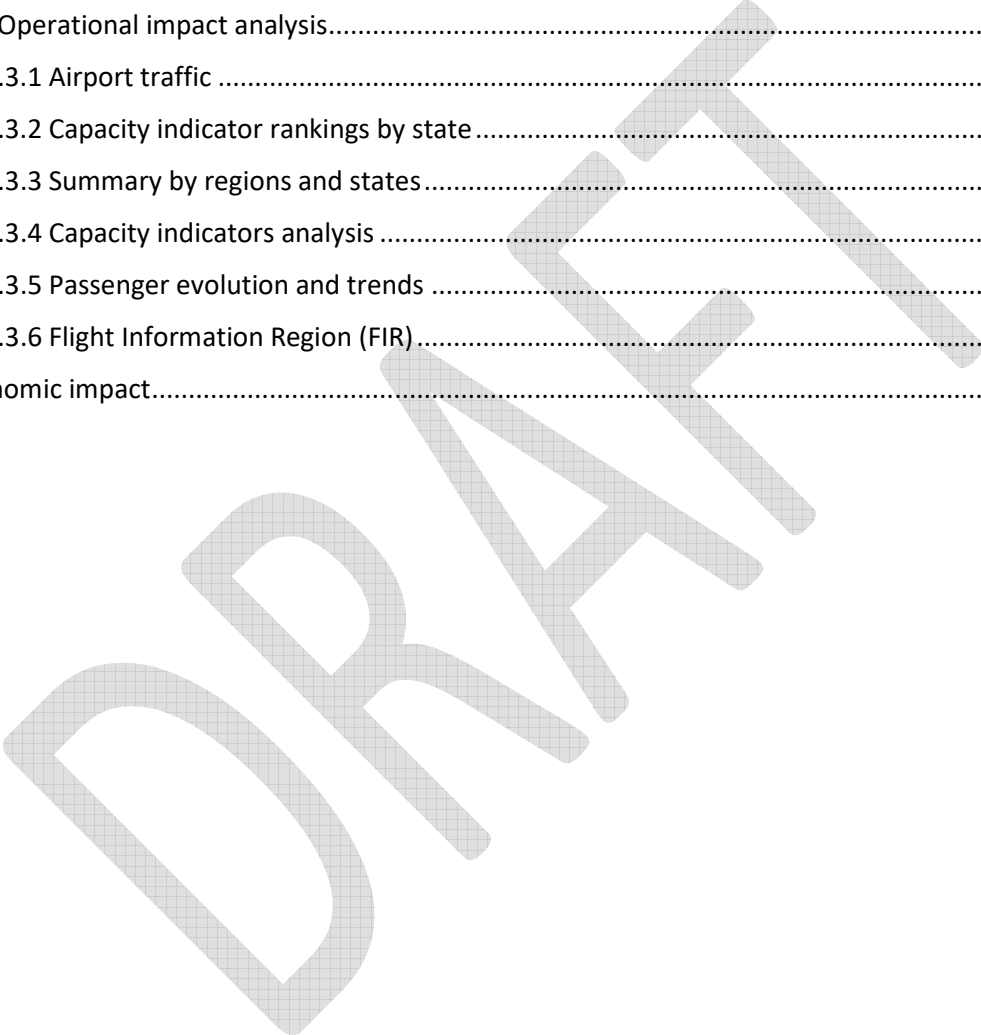
MIDT raw data is about 50 million rows per year. Combined with the Reference data, we use HIVE language in the Azure Cloud cluster to generate all the segments Origin and Destination of the MIDT raw data. Combined with the Reference data, we generate the MIDT monthly summary files in the Azure Cloud Storage Account. MIDT monthly summary files include information about Origin, Destination, Airline, Region, Route Group, Number of passengers. MIDT Monthly summary files are downloaded to ICAO network drive. MIDT Monthly summary files are loaded into the Business Intelligence platform (Tableau) which analyzes the data, creates the visualizations, and distributes it to the proper audience. The metadata of the MIDT data and the architecture of joining with the reference tables in EDM is available at [LINK].

# ICAO - AVIATION BIG DATA DASHBOARDS

Contents

1. Dashboard description

1.1 Dashboard categories

The COVID-19 Pandemic Traffic Data Analysis suite provides dashboards divided in 3 categories. The user can access dashboards in each section through a simple graphical interface as shown in Figure1. Detailed description of dashboard views within each section is provided below.

Figure 3 Dashboard category menu
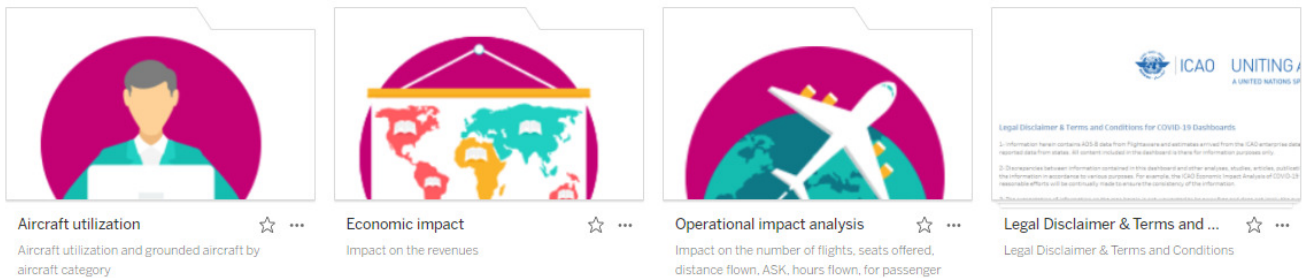
1.2 Aircraft utilization

This category includes Active fleet analysis by aircraft type dashboard view. This view provides a comparison of active fleet between a base and comparison year selected by the user. The data is shown by aircraft type[1] and month in a table format as well as bar charts.

1.3 Operational impact analysis

This dashboard category contains the following dashboard views and sections:

1.3.1 Airport traffic

This dashboard provides a comprehensive view of operational traffic data by origin and destination airport. The user selects the desired traffic indicator from a scroll down menu which includes number of flights, ASK flown and seats. Additional scroll down menus is provided to allow filtering results by airport,

---

[1] Regional jet, Single Aisle, Twin aisle, Cargo

country, routegroup, service type[2], Stage type[3] and time series. Selected results are illustrated in 3 different graphical formats as described below:

a) **Choropleth map:** an interactive map that is shaded according to aggregate country level value for selected traffic indicator. When user hover the pointer over each country a line graph with monthly time series values is shown for that country.

b) **Bar chart:** this chart shows sum of traffic for selected indicator by airport and sorts the results in desired order (ascending, descending)

c) **Pie chart:** this chart illustrates the proportional distribution of traffic for selected indicator by state

1.3.2 Capacity indicator rankings by state

This section contains 4 different dashboard views as described below:

a) **Operational indicators analysis by regions and states:** this view provides summary information for desired traffic indicator[4] selected by the user from a scroll down menu. Additionally, the user can filter the results by region, service type, and stage type and time series. Annual results by state are presented in a choropleth map and in table format on top of the page. Regional time series are presented in a line graph and in table format on the bottom. These results include useful summaries for selected indicators such as year on year percentage difference.

b) **Operational indicators analysis by Route Group:** this view provides similar dashboard with traffic indicators and filtering options at route group level.

c) **Ranking by States:** This view provides a dynamic ranking of states for desire indicator select by the user. Results can be filtered by stage type, service type and time series.

d) **Traffic Indicators Dashboard:** This view places all operational traffic indicators in the same page. State level aggregated values for each indicator is illustrated in a separate bar chart. Results can be filtered by region, route group, stage type, and service type and time series.

1.3.3 Summary by regions and states

This dashboard provides an overview _summary_ of global aircraft departures for desired time frame. Proportional distribution of departures by region and states is shown in pie charts. Additionally,

---

[2] Cargo/Passenger
[3] International/Domestic
[4] Number of flights, ASK flown, Distance flown, Hours flown, Seats

Departures by state are illustrated in choropleth map format and bar charts. The user can filter the results by desired region, stage type and time series.

1.3.4 Capacity indicators analysis

This section includes 6 dashboard views providing analysis of operational traffic indicators as well as *month on month* (MoM) analysis. User can filter results by region, country, stage type and service type. Dashboard views in this section are listed below:

a) **Departure analysis:** Monthly departures comparison for all available years.
b) **ASK analysis:** Monthly ASK comparison for all available years.
c) **Distance flown analysis:** Monthly distance flown comparison for all available years.
d) **Hours flown analysis:** Monthly hours flown comparison for all available years.
e) **Seats analysis:** Monthly seats flown comparison for all available years.
f) **MoM comparison:** this dashboard provides percentage month on month difference (growth/decline) for all operational indicators.

1.3.5 Passenger evolution and trends

This dashboard presents passenger numbers over time. Aggregate figures are illustrated in line and bar chart format on top of the page. Passengers by route group as well as month on month analysis are provided in the bottom of the page in bar chart and table format respectively.

1.3.6 Flight Information Region (FIR)

This dashboard presents number of aircraft movements and flown distances for each FIR.

2. Economic impact

This category includes 3 sections for Air carriers, airports, and Air navigation Service Providers (ANSPs). Each section includes 2 dashboard views illustrating impact of pandemic on revenues as described below:

a) **Revenue impact by regions:** This dashboard view shows revenue loss by region. Aggregate results are presented in choropleth map and table format on top of the page. Year on year analysis provided through an interactive bar chart. Detailed values and month over month analysis for all regions provided in table format at the bottom of the page.
b) **Revenue impact by route groups:** This dashboard view shows revenue loss by route groups. Aggregate results presented in bar charts on top of the page. Detailed values and month over month analysis for all route groups provided in table format at the bottom of the page.