

DATA WAREHOUSE PROJECT PART 1 - THE ACME-FLYING USE CASE

Group members: González, Nashly, Salat, Jonàs

Group 12

In the following document, we will explain our assumptions and justifications for the UML Diagram, the database structure and the materialized views for addressing the KPIs based on our database structure and data availability.

1. UML Diagram

The UML diagram represents a **star galaxy constellation schema** for the Flying Use Case, which consists of two fact tables **Flight_Fact** and **Maintenance_Fact**, and three dimension tables **Aircraft_Dim**, **Time_Dim** and **Reporteur_Dim**.

2. Database Structured

Each fact table has **foreign keys** to link to dimension tables but no primary key, as it's not needed for KPI calculations. All fact table attributes are **NOT NULL** to ensure reliable KPI results.

Each dimension table has a **surrogate primary key** for efficient joins and easier maintenance. **Time_Dim** has **NOT NULL** constraints for time-based KPIs like ADIS. In **Aircraft_Dim**, **AircraftRegistration**, **AircraftModel**, and **Manufacturer** are **NOT NULL**, with **AircraftRegistration** set as **UNIQUE** for data integrity. In **Reporteur_Dim**, **Airport** is **NOT NULL** to support KPI integrity.

3. Key Performance Indicators:

In **Flight_Fact**, **Flight Hours (FH)**, **Delays (DY)**, and **Cancellations (CN)** from AIMS are the core attributes used to calculate key Aircraft Utilization and LogBook metrics. These attributes enable calculations for Delay Rate (DYR), Cancellation Rate (CNR), Technical Dispatch Reliability (TDR), Average Delay Duration (ADD), and various Report Rates such as RRRh, RRRc, PRRh, PRRc, MRRh, and MRRc. **TO** is calculated by counting all the flights where the CN(cancelation) variable is 0 (false). For the delay, we have assumed that the values stored are always in the specified range, and otherwise is NULL. This way, we can count the number of delays simply by counting the variable, and it's easy to do this check when ingesting the data. We also assume that FH is 0 when the flight is cancelled.

In **Maintenance_Fact**, **Out of Service Days (ADOS)** from AIMS, along with **Pilot Logbook Count (PLBC)** and **Maintenance Logbook Count (MLBC)** from AMOS, are critical for calculating Scheduled Aircraft Days Out of Service (ADOSS), Unscheduled Aircraft Days Out of Service (ADOSU), Aircraft Days In Service (ADIS), and additional Report Rates, ensuring comprehensive coverage of maintenance and operational reporting metrics. For calculating ADOSS and ADOSU, we check whether the "scheduled" variable is 1 or 0.

For a complete breakdown of how each metric is derived, please see the appendix or the project folder documentation.

4. Materialized views:

We have decided to not include any materialized views. At first, we contemplated two main reasons to create MV: to pre query some parameters, like having filtered the delays in the specified interval, or having a parameter such as TO, that is used in a lot of queries also simplified. But since we finally decided the delays are always in the needed interval and the KPIs are not queried individually, but many in the same query (as specified in the project

statement), we have finally decided to not use any materialized views, as they would not give a significant enough reduction in the query timers or dimensions.

APPENDIX

This appendix outlines the definitions and calculation methods for aircraft utilization and logbook metrics. Refer to the "Queries" folder for detailed SQL calculations.

1. Aircraft utilization metrics

- 1.1. Flight Hours (FH):** We will obtain it from AIMS System by calculating the difference between actual arrival and actual departure considering only non-cancelled flights.
- 1.2. Flight Cycles (TO):** from AIMS System, we identified a take-off by a successful departure, meaning the flight is not cancelled and has a recorded actual departure time.
- 1.3. Aircraft Days Out of Service (ADOS):** We can derive ADOS from ADOSS and ADOSU by summing the days when the aircraft was undergoing any kind of maintenance, both scheduled and unscheduled.
- 1.4. Aircraft Days Out of Service Scheduled (ADOSS):** To calculate ADOSS, we select only the scheduled maintenance records and count the unique days the aircraft was in scheduled maintenance from AIMS System.
- 1.5. Aircraft Days Out of Service Unscheduled (ADOSU):** To calculate ADOSU, we select only the unscheduled maintenance records and count the unique days the aircraft was in unscheduled maintenance from AIMS System.
- 1.6. Aircraft Days In Service (ADIS):** We can calculate ADIS by subtracting ADOS from the total operational days for a given period.
- 1.7. Daily Utilization (DU):** We can derive DU by dividing flight hours (FH) by days in service (ADIS), where ADIS is the total days in the month minus days out of service (sum of ADOSS and ADOSU).
- 1.8. Daily Cycles (DC):** We can obtain DC by dividing take-offs (TOFF) by the days in service (ADIS), where ADIS is the total days in the month minus days out of service (sum of ADOSS and ADOSU).
- 1.9. Delay Rate (DYR):** We can derive DYR by dividing the number of delays (DY) by the number of take-offs (TOFF) and multiplying by 100.
- 1.10. Cancellation Rate (CNR):** We can obtain CNR by dividing the number of cancellations (CN) by the number of take-offs (TOFF) and multiplying by 100.
- 1.11. Technical Dispatch Reliability (TDR):** We can extract TDR by subtracting from 100% the proportion of take-offs with delay or cancellation (sum of DY and CN divided by TOFF) and multiplying by 100.

1.12. Average Delay Duration (ADD): We can obtain ADD by summing the total delay (DY) minutes and dividing by the number of delays (DY), then multiplying by 100.

2. LogBook metrics:

2.1. Report Rate per Hour (RRh): We can derive RRh by dividing the sum of maintenance logbook entries (MLBC) and pilot logbook entries (PLBC) by the flight hours (FH) and multiplying by 1000.

2.2. Report Rate per Cycle (RRc): We can obtain RRc by dividing the sum of maintenance logbook entries (MLBC) and pilot logbook entries (PLBC) by the number of take-offs (TOFF) and multiplying by 100.

2.3. PIREP Rate per Hour (PRRh): We can derive PRRh by dividing the total pilot entries (PLBC) by the flight hours (FH) and multiplying by 1000.

2.4. PIREP Rate per Cycle (PRRc): We can extract PRRc by multiplying the total pilot logbook entries (PLBC) by 100, then dividing by the total number of take-offs (TOFF).

2.5. MAREP Rate per Hour (MRRh): We can derive MRRh by dividing the total maintenance entries (MLBC) by the flight hours (FH) and multiplying by 1000.

2.6. MAREP Rate per Cycle (MRRc): We can obtain MRRc by dividing the total maintenance entries (MLBC) by the number of take-offs (TOFF) and multiplying by 100.