Project Assumptions

After extracting the base operational data by connecting to the operational databases AIMS and AMOS while also using the provided additional data sources (lookups), we proceeded to the creation of the dimensions and fact tables. Each dimension was modularized lest we should end up reprocessing each and every data flow in case of failure. Data quality was improved by removing duplicates and incomplete values using tUniqRow and tFilterRow, as well as ‘select distinct’ in the SQL queries.

Due to the fact that the database is not that big in size, we chose to perform the majority of the selections through complex SQL queries.

Performance was improved by parallelizing the flow and assigning more memory to the Logbook Reporting fact job and People dimension. In the Aircraft Utilization fact job, we tried to assign more memory but we encountered an error saying “Too small maximum heap.”, so it was left as is.

Regarding the reliability of our ETL process, recovery points were added in the Aircraft Utilization fact job and People dimension.

In reference to the business rule stating that in Flights table, departure and arrival airports must be those in the flightID (unless this flight has been diverted); a diverted flight is included in the safety type of Operational interruptions alongside four other categories, so it is impossible to know whether the reason for a “safety” type interruption was exclusively a diverting flight. In our SQL query we just made sure the origin and destination coincide with the departure and arrival airports.