

ASSUMPTIONS

We assume that the most important purpose of creating the multidimensional schema is to analyse the company domain by obtaining the KPIs. Because of this we remove all the rows containing aircrafts or maintenance people not available in the CSV files. Moreover, some of the business rules were not included if they were considered not significant for the purpose and lowering the performance of the ETL.

After analysis of performance metrics we consider the operation of connecting to the external source and extracting data as costly in terms of time consumptions. That's why the number of the table inputs steps have been reduced to the minimum.

We assume that all the information provided in the ACME flying use case is correct and all the data not matching the requirements is incorrect. Referring to this it was considered that in OperationInterruption table we can only allow events of kinds "Delay" and "Safety".

We obtain information about maintenance from tables: MaintenanceEvents and OperationInterruption.

In our ETL we didn't consider checking the table Maintenance (AIMS) and joining it with MaintenanceEvents (AMOS) because it's data is not needed for and KPI and, according to the provided business rules, each event stored there must exist in some of the events in AMOS included in corresponding time interval. Considering this the join necessary to ensure this would be really costly.

We respect all the fields length given in Target: Star-join schemas in reference to maximum value of measure. In meaning that for example maximum number of delays per day is 99 because field is type Number(99).

We assume that canceled flight can't be delayed or have any flight hours. Delay is computed as the difference between scheduled and actual arrival. We assume that delay or cancellation of flight corresponds to scheduled departure.

Mistake in schema

There is one mistake which we were aware of. It was caused by the complicated use-case and our misunderstanding of domain at the beginning stages. The issue was spotted at the very end of process creation and we were not able to fix it. The scheduled and unscheduled days out of service are stored incorrectly. If the plane is out of service few days in a row, cumulated number of those days is stored in the first day's row (it is not splitted for separate days).

Structure

Whole process is executed in a single Pentho transformation. We use "Wait" for tables to wait for the other dependent tables. We found using single transformation easier for managing parallel operations. Job checks the existence of output tables and CSV files. We assume that AIMS and AMOS data sources are available.