Project: Data Warehouse design

- ACME-Flying Use Case -

We computed four star schemas – two focusing the analysis of aircraft key performance indicators (**flights**, **maintenance_events**) and the other two concentrating on logbook key performance indicators (**work_orders**, **work_orders_MAREP_only**). The star schema with the fact table **flights** is the only one receiving data from "AIMS" (Air Information Management System). The other star schemata get their data from "AMOS" (Aircraft Maintenance Operation System).

We assume "Work Orders" and "Maintenance Events" as two disjunct sets of data. Due to this we created different star schemata.

For calculating the KPI's, you need one star schema per problem a)-d) (mostly): the schema of fact table **flights** for a), **maintenance_events** for b), **work_orders** for c), and **work_orders_MAREP_only** for d).

All star schemas can be linked by *aircraft_registration_code* since this measure is contained in all fact tables. This is needed since for the calculation of the KPI's of b), c) and d) there is information needed also from the star schema of the fact table **flights** (information for calculating KPI's FH and TO).

In creating the star schemata, we have observed the integrity constraints for summarizability (disjointment, completeness, compatibility). In this Data Warehouse, data is not changed (no WRITE), it is READ only.

The names of the measures in the fact tables and the attributes in the dimensions are adopted from the expressions used in the data sources.

In the fact table we put measures with a huge, growing number of entries (e. g. *flight_id*: ~300,000 p. a.) and in the dimensions those with less, completed entries like *aircraft_registation_code* (125 planes).

The attributes / measures with a * in front will be generated out of the existing data (e. g. **flight_time_(actual)* as the balance of *arrival_time_(actual)* minus *departure_time_(actual)*).

The fact tables as well as the dimensions include only the information which is necessary computing the KPI's. By choosing the right select statement, the values for the key performance indicators will be calculated.

We assume that the formulas for the key performance indicators (e.g. DYR, CNR etc.) are given. They were formulated somewhere before (between data and flow).

Materialized views are not necessary since within our star schemas we want to reach a minimal structural complexity (number of facts, dimensions, attributes) (parts are reused for answering new requirements) so that there would not be e. g. a reduction of indexes due to creating materialized views.

Meysam Zamani Forooshani
Sarah Stefanie Zorenc