

JUSTIFICACIÓ

- Hem decidit crear 4 taules de fet diferents, una que correspon a cada patró de consulta que pot interessar als usuaris. Les justificacions són a sota, però la idea principal és que tenir una taula de fet per a cada patró ens permetrà un disseny de cada taula de fets que optimitzi el temps de resposta a cadascuna de les consultes i resulti en la menor redundància (sense comprometre l'eficiència).
- Hem fet un diagrama tipus galaxia ja que, tot i tenir 4 taules de fet diferents, totes tenen dimensions d'anàlisi compartides.
- A continuació es presenten justificacions per les taules creades per a les dimensions i les taules de fets per a cada tipus de consulta
- Hem contrastat els requisits de cada consulta amb les dades a les quals teníem accés per comprobar que tot el que volem arribar a guardar es pugui calcular en l'etapa d'ETL.

DIMENSIONS D'ANÀLISI

- TEMPS: 3 dels 4 patrons de consulta tenen una dimensión d'anàlisi temporal. Tenim dades que requereixen una granularitat a nivel de dia, i dades amb granularitats a nivel de mes o any. Hem decidit crear una taula de diez i una taula de mesos i anys. Tot i haver de fer un join a l'hora d'accendir a la granularitat més baixa, permetem tenir totes dues granularitats a la base de dades i ens estalviem fer dues taules de dimensión temporal (que seria molt més costós en termes d'espai). L'SGCB està optimitzat per fer joins de manera ràida, per tant ralentir una mica la consulta A) a canvi de l'estalvi d'espai ens ha semblat un bon balanç.
- Aircrafts/models/manufacturers: No crearem diferents taules per a les diferents jerarquies d'agregació. L'únic cas en què aquesta estrategia tindria sentit és si el model o fabricant d'una aeronau específica fossin volàtils i canviéssin. Un avió no canviarà de model (ja s'ha construït com a un cert model, no canviarà) ni de fabricant (ja l'han fabricat). Per tant, com que són estatics, guardarem una sola taula amb els tres nivells de la jerarquía per a tots els fets que necessitin d'aquesta dimensión d'anàlisi.
- Aeroports: guardarem una sola taula amb els aeroports de treball del personal de manteniment per a les consultes D). No guardem informació sobre els treballadors, només els aeroports, perquè aquesta és la dimensión d'anàlisi que ens interessa, i guardar dades de persones a què no accedirem seria un malbaratament de l'espai. Ténim accés a les dades i podem construir la llista d'aeroports a l'etapa d'ETL.

CONSULTA A)

- guardem les hores de vol i nombre de vols diàries de cada aeronau. Dues dimensions d'accés, aeronau i dia

-La taula de fets guardarà el temps de vol i els departures totals per a cada dia. No guardarem dades amb menor granularitat perquè no les necessitem, i a més granularitat perquè ens impossibilitaria l'anàlisi que es demana.

CONSULTA B

- Guardem les estadístiques menusals de cada aeronau. Dues dimensions d'accés, aeronau i mes
- Guardem només ADOSSU i ADOSS perquè el càlcul de ADOS i ADIS a partir d'aquestes és una suma (súper ràpid de computació) . Podriem argumentar que com que é suna operación inmediata es poden guardar ja les mètriques i així ja ho tenim tot, però llavors augmentem el volum de dades de manera prou considerable i "per no res" (no ens estaviem gaire feina). Suposant que la taula final té (2 keys + 5 metriques) estariem afegint 2 mètriques (incrementem en quasi un 30% el volum de dades de la taula)
- PER DECIDIR: guardar nombres o ràtios de cancel·lacions, retards i averages

CONSULTA C

- Guardem la suma mensual de reports per a cada aeronau, separada per pilots i personal de manteniment. Les dimensions d'anàlisi són el mes i l'aeronau.
- El fet l'anomenem monthly report logs. No necessitem dades a més granularita que el nombre de logs per mes. S'afegeix la diferenciació entre pilots i manufacturers, perquè es requereix a les mètriques. Tanmateix NO guardem la informació de cada treballador (ja que no la necessitarem per res) sinó un agregat per cada mes i aeronau per cada rol de treballador (pilots i manteniment). L'única cosa que ens interessa és aquesta diferenciació i la guardem en dues mesures a la taula de fets.
- Tot i haver de mesurar ràtios amb TO l'agafarem de la taula de fet A) ja que és el càlcul d'un agregat mensual ràpid. D'aquesta manera donem més llibertat a l'usuari i simplifiquem les actualitzacions. **No guardarem els ràtios directament, PER QUÈ?**
-

CONSULTA D

- Guardem la suma TOTAL del nombre de reports fets per personal de manteniment per a cada aeronau a cada aeroport. Les dimensions d'anàlisi són l'aeroport i l'aeronau.
- Guardem els valors totals agregats i no una granularitat menor perquè no es necesita. Tot i que es podrà argumentar que afegim redundància (guardem quantitat de logs fets per personal de manteniment en dos llocs diferents), en aquesta taula (vs la C) afegim una dimensió d'anàlisi Aeroport, i guardem un agregat total directament en comptes de tenir una fila per cada mes, optimitzant així el temps de resposta a les consultes de tipus D).
- Guardem agregats i no els ràtios perquè aquí el TO es pot fer amb una aggregació de tota la columna TO de la taula daily info (A), serà ràpid i eficiente. A més a més, en actualitzar amb noves dades s'hauria de desfer el ràtio per poder recalcular l'agregat i afegir els valors nous. Considerem més adient fer el càlcul cada vegada. **Alguna justificació més?**

CONSIDERACIÓ SOBRE LES TAULES PER A LES CONSULTES C I D

- Hem pensat d'afrontar les opcions Ci D conjuntamente ja que comparteixen mètriques. Hem considerat dues taules de fets per guardar la informació de reports de pilots i manteniment per separat (perquè els pilots no tenen aeroport assignat i el manteniment sí). Ho hem descartat perquè comportaria guardar la mateixa informació de #reports sobre cada mes i aircraft en dues taules diferents. Això fa créixer tant el nombre de files total (el dupliquem, o més ja que afegim info. Aeroports), i en fer consultes del tipus c, més taules implica fer més joins (el doble) -> problemàtic.
- A més a més, en la consulta la dimensió d'aeroports és irrelevante
- I, tot i que guardem info. amb certa redundància, en la taula per a la consulta d) guardem agregats totals per aeroport i aircraft, cosa que farà molt més ràpida la query en el cas d). Si haguessim volgut no guardar els totals i calcular-los, com ja s'ha mencionat anteriorment s'haurien de guardar les dades en dues taules separades (s'indexen per dimensions diferents, segons el rol), ocuparia significativament més espai.