

Master 1 in ECONOMICS  
Master 1 ECONOMIE et STATISTIQUE  
**EPREUVE: DATA BASES – Code : M1-S1-12**
**Documents unauthorised (closed books)**
**Duration: 1h00**

A network of airline companies has built the following relational data warehouse to monitor travels:

**PASSENGER**(PassportNum, FNamePa, LNamePa, CityPa, CountryPa)

**IS\_TRAVELLING**(PassportNum#, CodeTr#)

**TRAVEL**(CodeTr, NameTr, Date, CodeAP\_Departure#, CodeAP\_Arrival#)

**AIRPORT**(CodeAP, NameAP, CityAP, CountryAP)

**FLIGHT**(FlightNum, Date, CodeAP\_Departure#, CodeAP#\_Arrival, CodeCo#, Distance, Time)

**COMPANY**(CodeCo, NameCo, CityCo, CountryCo)

**BOARDING\_PASS**(PassportNum#, FlightNum#, CodeTr#)

In this schema, primary keys are underlined and foreign keys are suffixed by the # sign.

This databases records passengers that travel from one airport to another. A travel may correspond to several flights (e.g. the travel from Toulouse to Montreal can be composed of the two flights Toulouse-Paris and then Paris-Montreal). Each passenger has a first and last name, is located in a city which is in a country. Airports and airline companies have a name and are located in a city which is in a country. Travels have a name, a start date, a departure airport and an arrival airport. Companies operate flights that have a date, a departure and an arrival airport, a distance (in Km) and the time taken by the flight (in hours). As a travel can correspond to several flights, passengers need one boarding pass for each flight they take and boarding passes are associated to travels.

**Work to be done (Question 1: 8 points; question 2: 8 points; question 3: 4 points)**

1. Company managers wish to obtain information about the travels done by passengers. Give the **SQL queries** that answer the following questions on the tables of the data warehouse (note that the Month and Year functions allow extracting the month and year number of a date):

1.1. Name of passengers that travel only with one company.

1.2. Name of the company countries, company name and the sum of the distances of all their flights only for companies that operated more than 1000 flights during 2015.

2. This data warehouse is used as a support for the construction of a **multidimensional data mart** for the manager of the airline network. This user wishes to analyse the distance, time and number of passengers of flights of companies, weekly, monthly and yearly and according to passenger cities. The non-temporal dimensions must have **all** the aggregation levels that can be extracted from the data warehouse, as well as **all** the associated information.

2.1. Provide the content of the **first three columns** of the measure dictionary of this data mart (columns of the dictionary are Code, Description and Type but not Extraction formula);

2.2. If there was an “Airport” dimension built using the Airport table (described above), what type of rule could be used to build this dimension’s hierarchy of attributes and why? Recall that there are three possible rules: general knowledge, business rules and functional dependencies (data that depend on other data). Justify the answer in one sentence.

2.3. Provide a graphical representation of the conceptual schema of this data mart (use the graphical formalism used during course exercises).

3. This data warehouse will be mainly used with a multidimensional data mart (i.e. an OLAP tool). Provide the answer for following questions:

3.1. Draw the schema of the general architecture of this decision support system.

3.2. In one sentence, state why multidimensional modelling is interesting for data analysis and decision support systems.

Master 1 in ECONOMICS  
Master 1 ECONOMIE et STATISTIQUE  
**EPREUVE: DATA BASES – Code : M1-S1-12**

**Documents non autorisés**

**Durée : 1h00**

Un réseau de compagnies aériennes a construit l'entrepôt de données relationnel suivant :

**PASSAGER**(NumPasseport, NomPa, PrenomPa, VillePa, PaysPa)

**VOYAGER**(NumPasseport#, CodeTr#)

**TRAJET**(CodeTr, NomTr, Date, CodeAR\_Depart#, CodeAR\_Arrivee#)

**AEROPORT**(CodeAR, NomAR, VilleAR, PaysAR)

**VOL**(NumVol, Date, CodeAR\_Depart#, CodeAR\_Arrivee, CodeCo#, Distance, Duree)

**COMPAGNIE**(CodeCo, NomCo, VilleCo, PaysCo)

**CARTE\_EMBARQUEMENT**(NumPasseport#, NumVol#, CodeTr#)

Dans ce schéma, les clés primaires sont soulignées et les clés étrangères sont suffixées par #.

Cette base de données enregistre les passagers qui voyagent d'un aéroport à un autre. Un trajet peut correspondre à plusieurs vols (par exemple : le trajet depuis Toulouse vers Montréal peut être composé de deux vols Toulouse-Paris suivi de Paris-Montréal). Chaque passager a un nom et un prénom, est situé dans une ville qui est dans un pays. Les aéroports et les compagnies aériennes ont un nom et sont localisés dans une ville qui est dans un pays. Les trajets ont un nom, une date de début, un aéroport de départ et un aéroport d'arrivée. Les compagnies opèrent des vols qui ont lieu à une date donnée, un aéroport de départ et un aéroport d'arrivée, une distance (en Km) ainsi que la durée du vol (en heures). Un trajet peut correspondre à plusieurs vols, les passagers ont besoin d'une carte d'embarquement pour chaque vol qu'ils empruntent. Les cartes d'embarquement sont associées à des trajets.

**Travail à réaliser (question 1 : 8 points ; question 2 : 8 points ; question 3 : 4 points)**

1. Les directeurs des compagnies souhaitent obtenir des informations sur les trajets effectués par les passagers. Donner le **code SQL des requêtes** suivantes sur les tables de l'entrepôt de données (remarque : les fonctions Month et Year permettent d'extraire le mois et l'année d'une date) :

1.1. Nom des passagers qui voyagent seulement avec une unique compagnie.

1.2. Nom des pays des compagnies, nom des compagnies et la somme des distances de tous leurs vols, uniquement pour les compagnies qui ont opéré plus de 1000 vols durant l'année 2015.

2. Cet entrepôt de données sert de support à l'élaboration d'un **magasin de données multidimensionnel** pour l'administrateur du réseau de compagnies aériennes. Cet utilisateur désire analyser la distance, la durée et le nombre de passagers des vols des compagnies, par semaine, par mois et année ainsi que selon les villes d'origine des passagers. Toutes les dimensions non temporelles doivent avoir **tous** les niveaux d'agrégation qui peuvent être extraits de l'entrepôt de données, ainsi que **toutes** les informations associées.

2.1. Donner le contenu des **trois premières colonnes** du dictionnaire des mesures de ce magasin de données (les colonnes de ce dictionnaire sont Code, Description, Type et Formule d'extraction) ;

2.2. S'il y avait eu une dimension « Aéroport » construite à partir de la table AEROPORT (décrite au début de ce sujet), quel type de règle pourrait être utilisée pour construire la hiérarchie des paramètres de cette dimension et pourquoi ? Pour rappel, il existe trois règles : connaissances générales, règle de gestion et dépendances fonctionnelles (dépendances entre les données). Justifier la réponse avec une phrase.

2.3. Donner une représentation graphique du schéma conceptuel de ce magasin de données (respecter les formalismes étudiés en cours).

3. Sachant que cet entrepôt de données sera principalement utilisé avec un magasin de données (c'est-à-dire un outil OLAP), répondre aux questions suivantes :

3.1. Donner le schéma de l'architecture générale de ce système d'aide à la prise de décision.

3.2. En une phrase, donner l'intérêt de la modélisation multidimensionnelle dans le cadre de l'analyse de données et de l'aide à la prise de décision.