

**Université Toulouse 1 Capitole
Ecole d'économie de Toulouse**

Année universitaire 2016-2017

Session 1

Semestre 2

Master 1 Economics, Econometrics & Statistics

Épreuve : Data Bases

Date de l'épreuve : 27 mars 2017

Durée de l'épreuve : 1h30

Liste des documents autorisés : Tous

List of authorized documents: All

Liste des matériels autorisés : Aucun

List of authorized equipment: None

Nombre de pages (y compris page de garde) : 7 – English instructions start on page 2

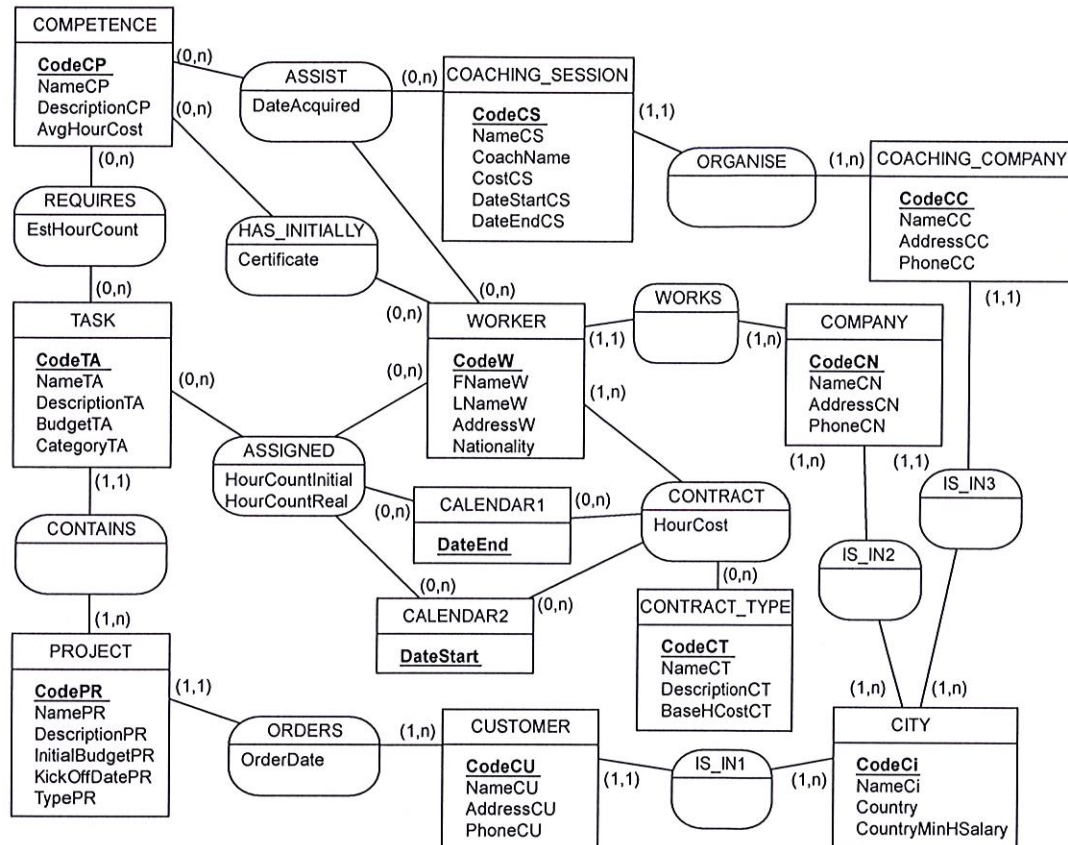
Number of pages (including this page) Instructions en français page 5

DATABASES – DECISION SUPPORT

This exam text is composed of 3 pages

All documents allowed (no electronic devices). –Length: 1h30–

A company that manages projects has assembled the following data warehouse. This data warehouse analyses the projects ordered by customers as well as the workers assigned to the project tasks and the associated competences required. The conceptual schema (entity/relationship formalism) is:



Each Project is ordered by a customer at a date and is dedicated an initial budget. The project will be launched at a kick-off date and is of a specific type (Product design, Software design, Web site, etc.). A project contains several tasks (budget management, case design, compliance testing, etc.) of a specific category (management, design, tests, etc.) and each task is given a budget. Tasks require an estimated number of hours of competences and each competence has an average hour cost. In order to do the projects, project tasks are assigned to workers for an initial number of hours a starting date and an end date (for example the project management task will be assigned to the worker who will be the project manager). The real number of hours spent on each task will be recorded once it is known. In order to do tasks, workers have competences either directly when hired (the name of the certificate validating the competence is recorded) or acquired at a specific date by assisting at coaching sessions. Each coaching session is done by a coach during one or more days, cost a certain amount of money and is organised by one coaching company.

Workers work for one company. They are hired for tasks by a contract with a starting and an ending dates. A worker may be hired several times for one or more project tasks or he may be hired only for part of the task. Each contract is based on one contract type (for example senior manager, junior designer, etc.) that indicates the base hour cost for the worker. Each worker is paid according to an hour cost which is negotiated with his company and greater or equal to the base cost. This hour cost

Exam – Session 1, March 2017

must also be, at least, greater or equal to the minimum hour salary of the country where the company is located.

The relational schema of the data warehouse is the following:

PROJECT (CodePR, NamePR, DescriptionPR, InitialBudgetPR, KickOffDatePR, TypePR)
ORDERS (CodePR#, CodeCU#, OrderDate)
CUSTOMER (CodeCU, NameCU, AddressCU, PhoneCU, CodeCI#)
TASK (CodeTA, NameTA, DescriptionTA, BudgetTA, CategoryTA, CodePR#)
REQUIRES (CodeTA#, CodeCP#, EstHourCount)
COMPETENCE (CodeCP, NameCP, DescriptionCP, AvgHourCost)
COACHING_SESSION (CodeCS, NameCS, CoachName, CostCS, DateStartCS, DateEndCS, CodeCC#)
COACHING_COMPANY (CodeCC, NameCC, AddressCC, PhoneCC, CodeCI#)
WORKER (CodeW, FNameW, LNameW, AddressW, Nationality, CodeCN#)
ASSIGNED (CodeTA#, CodeW#, DateStart, DateEnd, HourCountInitial, HourCountReal)
HAS_INITIALLY (CodeCP#, CodeW#, Certificate)
ASSIST (CodeCP#, CodeW#, CodeCS#, DateAcquired)
CONTRACT (CodeW#, CodeCT#, DateStart, DateEnd, HourCost)
CONTRACT_TYPE (CodeCT, NameCT, DescriptionCT, BaseHCostCT)
COMPANY (CodeCN, NameCN, AddressCN, PhoneCN, CodeCI#)
CITY (CodeCI, NameCI, Country, CountryMinHSalary)

In this schema, primary keys are underlined and foreign keys are followed by a hash (#).

Work to be done (3 exercises each with 3 questions)

1. Modelling a multidimensional data mart

The data warehouse is used to create a multidimensional data mart for the manager that supervises all projects. This manager wishes to analyse the monthly and yearly costs in terms of theoretical working hours, the real hours and the difference between these two costs (to see if there are extra charges). These costs have to be analysed according to the worker assigned, the task concerned and the month. All non temporal dimensions will use all aggregation levels that can be extracted from the data warehouse as well as all possible associated information. Note that the cost of task hours, is the total hour count spent by a worker multiplied by the hour cost of this worker.

- 1.1. Provide the content of the measure dictionary (the columns of this dictionary are: Code, Description, Type, Extraction formula). Remember that the description is very important.
- 1.2. Provide a complete graphical representation of the conceptual schema of this data mart (use the graphical formalism of facts, dimensions and hierarchies studied during the course).
- 1.3. For each hierarchy of the Task dimension, state which type of rule was used to design it.

2. SQL querying of the data warehouse

The training manager of the company wants to analyse data related to competences. Write the following SQL queries that will be executed on the tables of the data warehouse (note that the tables are described by the **relational schema** of the data warehouse).

- 2.1. For each coach, give the total number of sessions (we will suppose that there are no coaches with identical names);
- 2.2. Provide the list of workers (code, first and last names) that have the both competences “project management” and “budget simulation”;
- 2.3. In order to find experienced project managers, provide the code, the first and last names of each worker that has been assigned more than 5 times to tasks of the category “Project Management”.

3. Decision Support systems

Answer the following questions (other questions are on the next page):

- 3.1. In decision support systems, the basic tool is a spreadsheet. Give the reasons why this tool is widely used for visualising decisional data.

Exam – Session 1, March 2017

- 3.2. When building a data mart using a database tool, what are the three possibilities (or four depending on how you consider the architecture) available for implementing the system.
- 3.3. Describe briefly what one will find inside a Business Objects Universe.

Same Instructions in French follow.

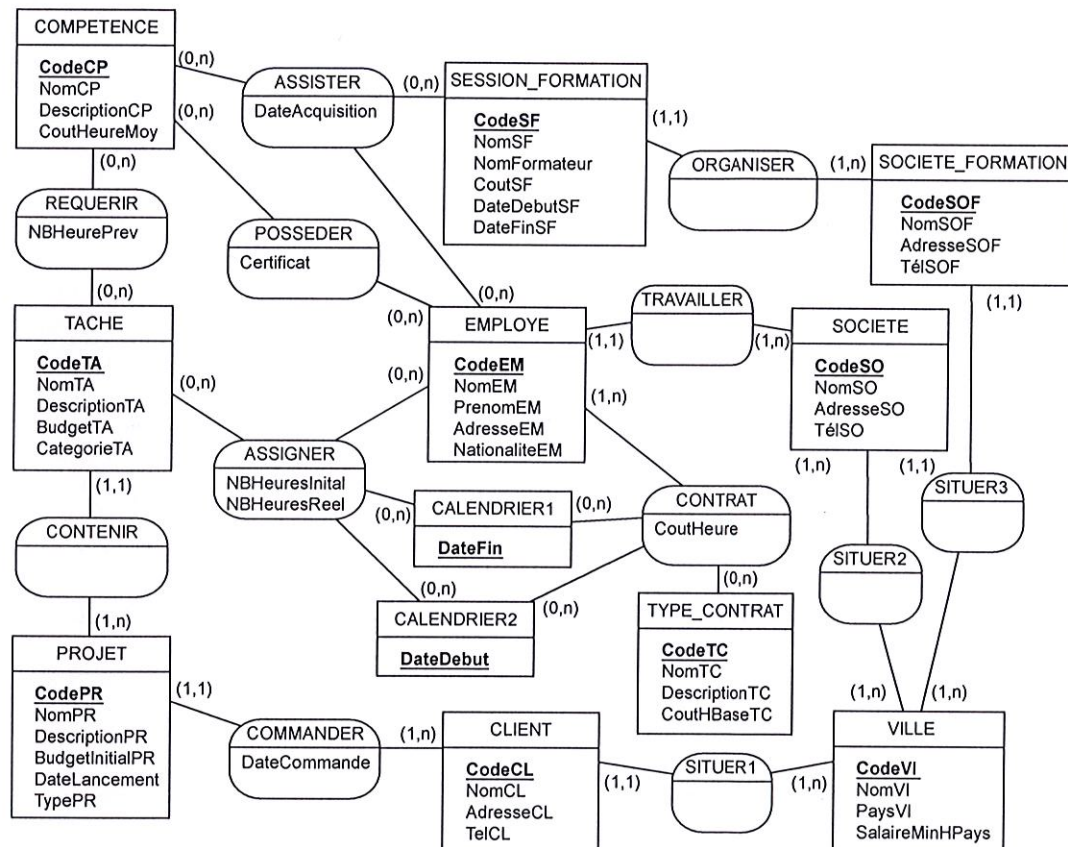
Instructions en français ci-après.

BASES DE DONNÉES – AIDE À LA DÉCISION

Ce sujet comporte 3 pages

Tous documents autorisés, aucun appareil électronique. –Durée : 1h30–

Une société qui gère la gestion de multiples projets a créé l'entrepôt de données suivant. Cet entrepôt permet l'analyse des projets commandés par des clients ainsi que les employés (provenant de sociétés sous-traitantes) assignés aux tâches du projet et les compétences requises. Le schéma conceptuel au formalisme entité/association est :



Chaque projet est commandé par un client à une date et un budget initial lui est dédié. Ce projet sera commencé à une date de lancement et est d'un type spécifique (conception de produit, conception de logiciel, Site Web...). Un projet contient un ensemble de tâches (gestion du budget, design de l'emballage, tests de conformité...) d'une catégorie spécifique (gestion, conception, tests...) et à chaque tâche est assignée un budget. Les tâches requièrent un nombre prévisionnel d'heures de compétences et chaque compétence a un coût moyen horaire. Afin de réaliser les projets, les tâches sont assignées à des employés pour un nombre initial d'heures, une date de début et une date de fin (par exemple, la tâche gestion du projet sera assignée au directeur du projet). Le nombre réel d'heure effectué pour chaque tâche est enregistré lorsqu'il est connu.

Afin de réaliser les tâches, les employés ont des compétences soit directement à leur embauche (le nom du certificat validant la compétence est alors enregistré) ou acquis à une date spécifique en assistant à une session de formation. Chaque session de formation est effectuée par un formateur, dure un ou plusieurs jours, coûte un certain montant d'argent et est organisée par une société de formation.

Les employés travaillent pour une société. Ils sont embauchés sur des tâches par un contrat avec une date début et de fin. Un employé peut être embauché plusieurs fois pour une ou plusieurs tâches d'un projet ou bien il peut être embauché que pour une partie d'une tâche.

Chaque contrat est basé sur un type de contrat (par exemple gestionnaire senior, concepteur junior...) qui indique le coût à l'heure de base pour l'employé. Chaque employé est payé selon un coût horaire qui est négocié avec sa société et qui est supérieur ou égale au coût de base. Ce coût horaire est également, au moins, égal au salaire horaire minimal du pays ou la société est localisée.

Le schéma relationnel de l'entrepôt de données est le suivant :

PROJET (CodePR, NomPR, DescriptionPR, BudgetInitialPR, DateLancementPR, TypePR)
COMMANDER (CodePR#, CodeCL#, DateCommande)
CLIENT (CodeCL, NomCL, AdresseCL, TelCL, CodeVI#)
TACHE (CodeTA, NomTA, DescriptionTA, BudgetTA, CategorieTA, CodePR#)
REQUERIR (CodeTA#, CodeCP#, NBHeurePrev)
COMPETENCE (CodeCP, NomCP, DescriptionCP, CoutHeureMoy)
SESSION_FORMATION (CodeSF, NomSF, NomFormateur, CoutSF, DateDebutSF, DateFinSF, CodeSOF#)
SOCIETE_FORMATION (CodeSOF, NomSOF, AdresseSOF, TelSOF, CodeVI#)
EMPLOYE (CodeEM, NomEM, PrenomEM, AdresseEM, NationalitéEM, CodeSO#)
ASSIGNER (CodeTA#, CodeEM#, DateDebut, DateFin, NBHeuresInitial, NBHeuresReel)
POSSEDER (CodeCP#, CodeEM#, Certificat)
ASSISTER (CodeCP#, CodeEM#, CodeSF#, DateAcquisition)
CONTRAT (CodeEM#, CodeTC#, DateDebut, DateFin, CoutHeure)
TYPE_CONTRAT (CodeTC, NomTC, DescriptionTC, CoutHBaseTC)
SOCIETE (CodeSO, NomSO, AdresseSO, TelSO, CodeVI#)
VILLE (CodeVI, NomVI, PaysVI, SalaireMinHPays)

Dans ce schéma, les clés primaires sont soulignées et les clés étrangères sont suffixés d'un #.

Travail à réaliser (3 exercices avec chacun 3 questions)

4. Modélisation d'un Magasin de Données multidimensionnel

L'entrepôt de données est utilisé pour créer un magasin de données multidimensionnelles pour le manager qui supervise l'ensemble des projets. Ce manager, désire analyser mensuellement et annuellement le coût en terme d'heures travaillées initiales et réelles ainsi que de différence entre ces deux coûts (révélant les éventuels surcoûts). Ces coûts doivent être analysés selon l'employé affecté, la tâche concernée et le mois. Toutes les dimensions non temporelles devront utiliser tous les niveaux d'agrégation qu'il est possible d'extraire de l'entrepôt de données ainsi que toutes les informations complémentaires disponibles. Remarque : le coût des heures des tâches est le nombre total d'heures faites par un employé multiplié par le coût horaire de cet employé.

- 4.1. Fournir le contenu du dictionnaire des mesures (les colonnes de ce dictionnaire sont : Code, Désignation, Type, Formule d'extraction). Rappel : la désignation est la colonne la plus importante.
- 4.2. Dessiner le schéma multidimensionnel complet en respectant le formalisme vu en cours à base de faits, de dimensions et de hiérarchies.
- 4.3. Pour chaque hiérarchie de la dimension Tâche, citer la ou les règles qui permettent sa création.

5. Interrogation SQL de l'entrepôt de données

Le gestionnaire des formations de la société désire de son côté analyser les données liées aux compétences. Écrire les requêtes SQL suivantes qui seront exécutées sur les tables de l'entrepôt de données (remarque : ces tables sont décrites dans le **schéma relationnel** détaillé ci-dessus).

- 5.1. Pour chaque formateur, fournir le nombre total de sessions (nous supposons qu'il n'existe pas deux formateurs avec le même nom) ;
- 5.2. Fournir la liste des employés (code, nom et prénom) qui ont à la fois la compétence « Gestion de projet » et « Simulation de budget » ;
- 5.3. Afin de trouver des chefs de projets expérimentés, fournir le code, le prénom et le nom de chaque employé ayant été assigné plus de 5 fois à une tâche de la catégorie « Gestion de projet ».

6. Système d'aide à la prise de décision

Répondre aux questions suivantes :

- 6.1. Dans les systèmes d'aide à la prise de décision, l'outil de base est le tableur. Donner les raisons expliquant pourquoi cet outil est autant utilisé pour la visualisation de données décisionnelles.
- 6.2. Lors de la construction d'un magasin de données, quelles sont les trois possibilités (ou quatre en fonction de comment vous considérez l'architecture) disponibles pour implanter ce système.
- 6.3. Décrire brièvement ce que l'on trouve à l'intérieur d'un Univers Business Objects.