

Informatique décisionnelle et big data

ECTS
6 crédits

Composante
**Sciences Fondamentales
et Appliquées**

Période de l'année
Semestre 8

Présentation

Description

Entrepôt de données : modélisation multidimensionnelle, représentation logique (ROLAP, MOLAP, HOLAP), processus ETL, requête OLAP.

Optimisation de requêtes : plan d'exécution, modèle de coût, algorithmes de jointures, indexation, vue matérialisée, partitionnement, stockage en colonne.

Techniques pour le traitement des big data : bases de données distribuées, framework de type MapReduce.

Introduction à la fouille de données : règle d'association, recommandation, clustering, arbre de décision.

Objectifs

Cet UE propose une introduction à l'informatique décisionnelle et aux défis liés à la gestion des données massives (big data). Le but de ce module est d'être capable de mettre en place un système décisionnel à des fins d'analyse tout en répondant au besoin de performance des requêtes décisionnelles dans le contexte des données massives.

Heures d'enseignement

CM	CM	12h
TD	TD	18h
TP	TP	20h

Pré-requis obligatoires

Avoir déjà suivi un cours d'introduction aux bases de données (conception, modèle relationnel, normalisation, SQL).

Connaître les grands principes du fonctionnement d'un système de gestion de bases de données (architecture, organisation physique, utilisation de la mémoire centrale).

Programme détaillé

Les principales compétences développées sont les suivantes.

- * Mettre en place un système décisionnel permettant de regrouper les données d'une entreprise à des fins d'analyse.
 - Comprendre la différence entre un entrepôt et un lac de données.
 - Concevoir et implémenter un entrepôt de données.
 - Définir le processus ETL (Extraction, Transformation, Chargement) permettant de peupler un entrepôt de données.
- * Exploiter un système décisionnel
 - Définir des requêtes OLAP pour interroger un entrepôt de données.
 - Faire du reporting afin d'afficher les résultats de requêtes OLAP sous une forme adaptée.
 - Utiliser différents algorithmes de fouilles de données pour répondre à différents besoins d'analyse.
- * Répondre au besoin de performance des requêtes décisionnelles
 - Analyser un plan d'exécution d'une requête.
 - Définir des structures d'optimisation adaptées (index, vue matérialisée, partitionnement) pour optimiser une ou plusieurs requêtes.
 - Comprendre les avantages et inconvénients des systèmes de gestion de bases de données orientés colonnes
 - Mettre en place des solutions distribuées (Bases de données distribuées, framework MapReduce) pour traiter des données massives.

Compétences visées

Les principales compétences développées sont les suivantes.

- * Mettre en place un système décisionnel permettant de regrouper les données d'une entreprise à des fins d'analyse.
 - Concevoir et implémenter un entrepôt de données.
 - Définir le processus ETL (Extraction, Transformation, Chargement) permettant de peupler un entrepôt de données.
- * Exploiter un système décisionnel
 - Définir des requêtes OLAP pour interroger un entrepôt de données.
 - Faire du reporting afin d'afficher les résultats de requêtes OLAP sous une forme adaptée.
 - Utiliser différents algorithmes de fouilles de données pour répondre à différents besoins d'analyse.
- * Répondre au besoin de performance des requêtes décisionnelles
 - Analyser un plan d'exécution d'une requête.
 - Définir des structures d'optimisation adaptées (index, vue matérialisée, partitionnement) pour optimiser une ou plusieurs requêtes.
 - Mettre en place des solutions distribuées (Bases de données distribuées, framework MapReduce) pour traiter des données massives.