

Entrepôt de données & Datamining

Volume horaire : 28h

Objectifs pédagogiques

- Comprendre et être capable de mettre en œuvre les outils du datawarehouse
- Comprendre et être capable de mettre en œuvre les outils du datamining
- Être capable de choisir de manière justifiée le bon outil du datamining en fonction du problème rencontré
- Comprendre les points forts et faibles ainsi que les limites de ces outils

Programme

1. Entrepôts de données

- Notion d'entrepôts de données et de magasins de données
- Architecture d'un entrepôt de données
- Modélisation multi dimensionnelle des données
- Modélisation conceptuelle (Modèle en étoile, en flocon, en constellation)
- Notions d'administration : gestion des métadonnées & réplication
- Comparatif des solutions existantes (Oracle, IBM, SQLServer, Teradata)
- TP : Installation d'un entrepôt et administration

2. Alimentation de l'entrepôt

- Processus ETL : Extracting, Transforming, Loading
- Comparaison des outils d'ETL
- TP : Application d'un outil d'ETL

3. Analyse multidimensionnelle & reporting

- Comparaison des approches OLAP et OLTP
- Représentation des données
- Présentation des opérateurs OLAP
- Tableaux de bords et KPI (Key Performance Indicators)
- TP : Présentation d'un outil de reporting

4. Datamining et analyse des données

- Introduction au datamining, présentation des domaines d'application à l'aide d'exemples
- Exposition des différentes phases de traitement des données : échantillonnage, préparation des données, fouille des données, tests et validation
- Exposition des quatre modes d'apprentissage et notion d'apprentissage
- Comparaison entre problèmes d'apprentissage supervisé et non supervisé
- Introduction aux SVM (Support Vector Machines) et réseaux de neurones
- TP final : fouille de données textuelles.

Moyens et outils pédagogique

Cours théorique & TP de mise en situation

Évaluation

Devoir sur table & Évaluation du TP final

Bibliographie

- *Handbook of Statistical Analysis & Data Mining Applications* - **Robert Nisbet, John Elder et Gary Miner**
- *Data Mining et statistique décisionnelle* - **Stéphane Tufféry**
- *Entrepôts de données : guide pratique de modélisation dimensionnelle* - **Kimball R. et Ross M.**
- *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming and Delivering Data* - **R. Kimball et J. Caserta**