

Méthodes d'accès aux données

Domaine	Ingénierie et Architecture
Filière	Informatique et systèmes de communication
Orientation	Informatique logicielle (ISCL)
Mode de formation	Plein temps

Informations générales

Nom	: Méthodes d'accès aux données
Identifiant	: MAC
Années académiques	: 2021-2022, 2022-2023
Responsable	: Nastaran Fatemi
Charge de travail	: 120 heures d'études
Périodes encadrées	: 64 (= 48 heures)

Semestre	E1	S1	S2	E2	S3	S4	E3	S5	S6
Cours						32			
Laboratoire						32			

Connaissances préalables recommandées

L'étudiant-e doit connaître et savoir utiliser les notions suivantes :

- concepts de base du modèle relationnel (modélisation, langage SQL, formes normales) ;
- algorithmes et structures de données (graphes, tables, B-Arbres) ;
- langage Java (éléments de base du langage).

Les unités d'enseignement BDR (bases de données relationnelles), ASD (algorithmes et structures de données) et POO (programmation orientée objet, Java) permettent d'acquérir ces connaissances.

Objectifs

A l'issue de cette unité d'enseignement, l'étudiant-e sera capable de :

- décrire la problématique liée à l'adaptation du modèle Objet au modèle Relationnel (« Impedance mismatch ») ;
- maîtriser les différentes catégories de modèles de données NoSQL ;
- utiliser les langages de requêtes NoSQL ;
- décrire les principaux modèles de recherche d'information (RI) ;
- utiliser le modèle vectoriel pour l'indexation et la recherche d'information textuelle ;
- utiliser les modèles récents d'analyse de données "big data" tel que le map-reduce et le RDD de Spark.

A l'issue des travaux pratiques en laboratoire, l'étudiant-e sera en outre capable de :

- réaliser une application interagissant avec une base de données NoSQL ;
- réaliser une application d'indexation et de recherche des données textuelles via la librairie Apache Lucene. ;
- réaliser une application d'analyse de données big data avec map-reduce ou Apache Spark.

Contenu et formes d'enseignement

Répartition des périodes indiquée à titre informatif.

Cours: 32 périodes

- Introduction 2
- Principes de systèmes distribués et le théorème CAP 2
- Modèle de données et langages de requêtes NoSQL 8
- Indexation et recherche d'information textuelle avec le modèle vectoriel 8
- Analyse des données big data avec Spark 8
- Travaux écrits 4

Laboratoire: 32 périodes

- Réalisation d'une application de base de données avec une base de données NoSQL 12
- Indexation et recherche d'information vectorielle avec la librairie Apache Lucene 10
- Analyse de données big data avec Spark 10

Bibliographie

1. "Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems", Martin Kleppmann, Mars 2017.
2. "Next Generation Databases NoSQL and Big Data", Guy Harrison, 2015.
3. "Database Systems, The Complete Book", 2nd Edition, Hector Garcia-Molina, Jeffrey D. Ullman, and Jennifer Widom, 2008.

Contrôle de connaissances

Cours : l'acquisition des matières de cet enseignement sera contrôlée au fur et à mesure par des tests et des travaux personnels tout au long de son déroulement. Il y aura au moins 3 tests d'une durée totale d'au moins 2 périodes.

Laboratoire : ils seront évalués sur la base des rapports de manipulation, à 3 reprises au minimum.

Calcul de la note finale

Note finale = moyenne cours x 0.67 + moyenne laboratoire x 0.33