

INTITULE DE LA MATIERE : **BASES DE DONNEES NO SQL ET INDEXATION**

CODE : UE-74

SEMESTRE : 7

NOM DE L'ENSEIGNANT RESPONSABLE : Patrice Bellot

NOMBRE D'HEURES ENCADREES CM : 12 TD : 8 TP : 20

NOMBRE D'HEURES DE TRAVAIL PERSONNEL DE L'ELEVE – estimation : 25

OBJECTIF GENERAL (pas plus de 5 lignes) :

Ce cours s'intéresse à la question de l'entreposage et de l'accès aux données dans une perspective Big Data et accès et calcul distribués. Les principaux types de bases de données NoSQL sont présentés ainsi que les principaux environnement pour l'exploitation des données entreposées. Ils sont mis en perspective d'applications réelles.

CONTENU – PLAN DU COURS (pas plus de 15 lignes) :

- Présentation des différents types de bases NoSQL (clés-valeurs, colonnes, graphes, documents, orientées séries temporelles) et comparaisons avec des bases relationnelles
- Présentation des principales solution du marché (écosystème Hadoop, bases MongoDB, Cassandra, ElasticSearch, InfluxDB... mais aussi les moteurs de traitement Apache Spark et Talend)
- Indexation et requêtage des bases de données dans une perspective d'analyse de données, de recherche d'information ou de visualisation
- La question de la représentation et de l'indexation de données textuelles ou orales
- Méthodes de sélection statistique de données (filtrage du bruit) et de descripteurs, détection de signaux faibles
- Expérimentation avec une bases de données NoSQL notamment pour la détection d'anomalie, la prévision de la demande et l'analyse de messages texte (ElasticSearch) ou l'analyse temps réel (influxDB).

FORME DE L'EVALUATION :

Contrôle continu     Contrôle terminal     Mémoire/rapport     Soutenance  
préciser si nécessaire, nombre d'épreuves :                      et type oral / écrit

ACQUIS DE LA FORMATION ATTENDUS, ET QUI SONT EVALUES (5 à 10 items)

A l'issue de cet enseignement, l'élève est capable de :

*Utilisez les termes comprendre, analyser, exploiter, interpréter, calculer, utiliser, mettre en œuvre, choisir, communiquer, rédiger,...*

- Modéliser des gisements de données et des flux informationnels et les mettre en forme
- Déterminer quel type de base de données est le plus adapté aux besoins et aux données disponibles et quelles données doivent être stockées ou filtrées
- Exploiter les données recueillies avec des approches d'analyse de données vues au semestre précédent
- Expérimenter et conduire des analyses avec un environnement NoSQL

PREREQUIS (pas plus de 5 lignes): Cours Analyse de données, Java ou Python

REFERENCES, BIBLIOGRAPHIE (pas plus de 5 lignes) :

- Les bases de données NoSQL et le Big Data: Comprendre et mettre en œuvre, R. Bruchez, Dunod, 2015.
- Elasticsearch: The Definitive Guide: A Distributed Real-Time Search and Analytics Engine, C. Gormley, Z. Tong, O'Reilly, 2015
- Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems, M. Kippmann, O'Reilly, 2017

- NoSQL: Database for Storage and Retrieval of Data in Cloud, G. Deka, Chapman & Hall, 2017