

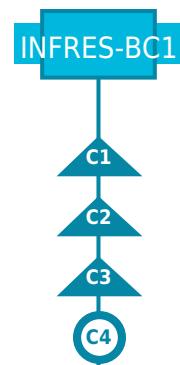
Pourquoi cette UE ?

Ce module répond au besoin de former les élèves aux environnements de données massives et hétérogènes. Le Big Data introduit les concepts, architectures et outils pour traiter de grands volumes distribués, tandis que NoSQL apporte des solutions de stockage flexibles et adaptées aux données non structurées. Ensemble, ces cours offrent une vision cohérente des technologies modernes de gestion des données et préparent à concevoir des systèmes performants et évolutifs.

Eléments constitutifs de l'UE

	coefficients	
INFRES_9_4_DL-1 Base de données NoSQL	1	
Volume d'heures d'enseignement encadré	Volume d'heures de travail personnel	Nombre d'ECTS
50	10	2

Alignement curriculaire

Parmi les compétences visées par la formation, lesquelles sont développées dans cette UE ?

- BC1 L'UE ne contribue pas à ce bloc de compétences
- BC1 L'UE contribue à ce bloc de compétences
- C1 Compétence non adressée dans cette UE
- C1 Compétence mise en œuvre dans cette UE
- C1 Compétence enseignée dans cette UE
- C1 Compétence évaluée dans cette UE
- C1 Compétence enseignée et évaluée dans cette UE

Contexte et enjeux de l'enseignement

Ce cours aborde les alternatives aux bases de données relationnelles pour la gestion des très hautes volumétries. Nommées NoSQL, ces nouvelles approches permettent de s'adapter aux problématiques liées aux nouveaux besoins en gestion de données induites par les géants du Web. L'objectif de ce cours est de brosser les différentes technologies disponibles afin de pouvoir choisir en toute connaissance la bonne technologie selon le besoin du projet.

Prise en compte des dimensions socio-environnementales

ODD9 - Industrie, innovation et infrastructure

Prérequis

Une bonne connaissance des bases de données relationnelles est nécessaire, ainsi que des notions d'architecture réseau et système d'exploitation.

Modalités d'enseignement et d'évaluation

	Nb d'heures
Cours	
Cours intégré (cours + TD)	12
TD	
TP	15
Projets	
Travail en autonomie encadré	
Contrôles et soutenances	1
Travail personnel	5

Objectifs pédagogiques

(à la fin de cet enseignement, l'étudiant sera capable de ...)

Les apprentis devront connaître les différentes typologies de bases NoSQL, ainsi que les fondements théoriques inhérents à ces bases. Cela leur permettra de choisir la base la plus en adéquation avec les problématiques qu'ils auront à traiter, chaque base ayant des caractéristiques fortement nuancées. Par ailleurs, les apprentis devront être capables, face à un cas concret, de mettre en place différentes solutions, un plan d'expérimentation rigoureux, afin de comparer ces différentes solutions et d'être à même de choisir la meilleure compte tenu des enjeux du projet.

Activités

(CM, TD, TP, projet, sortie terrain, etc.)

Cours théorique : 12h
TP sur machines personnelles (installation, configuration, implémentation des 2 solutions, tests, documentation) : 15h, travail effectué en binôme

Évaluations et retours faits aux élèves

(évaluations qui comptent pour la note ou qui permettent à l'étudiant de se situer, corrigés, feedback personnalisé...)

Evaluation :

- Rapport de TP (par binôme) type « professionnel » où les apprentis devront établir leur plan d'expérimentation, les tests réalisés / performances, et les conclusions et recommandations par rapport à la problématique originelle.
- QCM individuel permettant de vérifier l'acquisition des savoirs fondamentaux

Retour sur l'évaluation fait à l'élève :
Corrections du rapport et du QCM consultables sur demande avec débriefing individuel sur demande

Plan de cours

Le cours est divisé en 2 parties :

- Présentation théorique des bases NoSQL
 - o Limites des SGBDR, modèle relationnel vs agrégat, théorème de Brewer (CAP)
 - o Fondements des systèmes NoSQL : shardind, hashing, Map Reduce, MVCC, ...
 - o Modèles NoSQL : Clé-Valeur, Colonne, Document, Graphe
- Application et étude pratique de la base « Graphe » Neo4j :
 - o Concepts des bases « Graphe » (propriétés, noeuds, relations)
 - o Installation / configuration Neo4j
 - o Langage CYPHER

Le TP est constitué d'une étude de cas où les apprentis devront implémenter la problématique avec un SGBDR et une base NoSQL pour comparer les performances, avantages et inconvénients de chacune des solutions.

Ressources et références

Les supports pédagogiques sont disponibles en ligne sous Campus.

Contexte et enjeux de l'enseignement

Les Big Data (données massives) offrent des perspectives stimulantes dans de nombreux domaines industriels et académiques. Elles permettent notamment d'aborder de larges gammes de problèmes par une approche gouvernée par les données (data-driven). Cette approche peut s'avérer particulièrement intéressante pour découvrir des caractéristiques utiles pour diverses définitions stratégiques. Mais les Big Data induisent des challenges importants pour le stockage, transfert, et traitement de données ; elles nécessitent de repenser certaines approches.

Prise en compte des dimensions socio-environnementales

ODD9 - Industrie, innovation et infrastructure

Prérequis

Connaissances de 3ème année d'école d'Ingénieur par alternance, spécialité Informatique. Langage de programmation Java. Connaissance des bases de données relationnelles.

Modalités d'enseignement et d'évaluation

	Nb d'heures
Cours	
Cours intégré (cours + TD)	18
TD	
TP	3
Projets	
Travail en autonomie encadré	
Contrôles et soutenances	1
Travail personnel	5

Objectifs pédagogiques

(à la fin de cet enseignement, l'étudiant sera capable de ...)

Maitrise basique du concept de Données massives (Big Data), compréhension générale des enjeux associés et de l'intérêt potentiel. Compréhension des challenges techniques relatifs au traitement de données volumineuses et développement de capacités de mise en œuvre technique via l'outil Hadoop.

Activités

(CM, TD, TP, projet, sortie terrain, etc.)

Un cours magistral de 12h suivi de 12h de TP. La matière est dispensée via des séances de cours traditionnelles dans lesquelles les interactions enseignant-étudiant et entre étudiants sont nombreuses (e.g., échange des problématiques rencontrées par les étudiants en entreprise). Les séances pratiques sont centrées sur la résolution d'exercices nécessitant une analyse du problème proposé, une expression formelle de la solution, et une mise en œuvre technique au travers de la programmation de cette dernière. Les exercices sont généralement réalisés en binôme.

Évaluations et retours faits aux élèves

(évaluations qui comptent pour la note ou qui permettent à l'étudiant de se situer, corrigés, feedback personnalisé...)

Evaluation :
Contrôle et projet

Retour sur l'évaluation fait à l'élève :
Copies corrigés et consultables sur demande

Plan de cours

Présentation générale des notions et introduction des différentes dimensions des données massives sous la forme d'outils théoriques permettant de caractériser une problématique de ce type. Présentation des difficultés théoriques et pratiques induites par les Big Data. Nombreux exemples d'application et illustrations d'outils permettant l'analyse de ce type de données. Introduction au paradigme MapReduce et à l'outil Hadoop pour le traitement de données volumineuses.

Ressources et références

Les supports pédagogiques sont disponibles en ligne sous Campus.