



Guide pédagogique
Module « Collecte et Stockage de Données »
Options IASD et IL – 8.1 (4 crédits ECTS)

Place du module et enjeux

Poser les bases de la gestion des données, depuis leur acquisition (plan d'expérience, collecte) jusqu'à leur stockage (bases de données) et leur utilisation. L'objectif est de sensibiliser les étudiants à l'importance de l'organisation de la collecte des données, de leur qualité et de leur organisation au sein de bases de données en fonction des traitements qui vont les utiliser : problèmes de taille d'échantillons, de données imprécises voir manquantes, de redondance d'information, d'accessibilité aux données, etc.

Teaching guide and syllabus
« Data Collection and Storage » module
DSAI and SE options – 8.1 (4 ECTS credits)

Subject matter importance and associated issues

Lay the foundations of Data management, from their gathering (Design Of Experiment, collection) to their storage. The aim is here to make students aware of DOE's setup, data quality and arrangement within databases in regards to their future treatments: samples sizes, imprecise or missing data, redundancy, data accessibility, etc.

Responsable : Nicolas Sutton-Charani

Téléphone : 04 34 24 62 67

Courriel : nicolas.sutton-charani@mines-ales.fr



IMT Mines Alès
École Mines-Télécom

ENSEIGNEMENTS ACADÉMIQUES	Volume horaire	Détail des coefficients	Crédits
Collecte et Stockage de Données	50 h		
○ Collecte des données et plan d'expériences	18	1.5	4
○ Bases de données avancées	32	1.5	

Matière 1 :

Titre de la matière : Collecte des données et plan d'expériences	
Code : 2IA-8.1.1	Titre du module : « Collecte et Stockage de Données »
Semestre : S8	Cursus de rattachement : département 2IA

Heures présentiel	Heures total	Cours	TD	TP	Projet	Contrôles	Travail personnel	Coef /module	ECTS
18	38	6	0	5	6.25	0.75	20	1.5	2

Titre	Plan d'expérience et collecte des données
Résumé	Ce cours a pour objectif de poser les bases de la collecte de données. Depuis le plan d'expérience jusqu'à la collecte physique (questionnaires) ou virtuelle (Web) des données. Il y est en particulier question de qualité des données (imprécises, manquantes, aberrantes). La mise en application de ces enseignements dans le projet qui sera proposé dans la suite du module permettra de s'assurer que les étudiants les maîtrisent.

Responsable	Nicolas Sutton Charani – CERIS/IMT Mines Alès
Équipe enseignante	Nicolas Sutton Charani – CERIS/IMT Mines Alès Andon Tchechmedjiev – CERIS/IMT Mines Alès

Mots-clés	Plan d'expérience, collecte de données, Web scrapping, qualité des données,
Prérequis	Régression linéaire, Python. Les étudiants seront amenés à utiliser leur ordinateur personnel et doivent donc le prendre à chaque séance.

Contexte et objectif général :

Les enseignements des options IASD (Intelligence Artificielle et Science des Données) et IL (Ingénierie Logicielle) manipulent des données stockées de façon structurée ou non. La qualité des traitements informatiques sur ces données dépend de la qualité de ces données. Cet enseignement propose les bonnes pratiques pour assurer cette qualité.

Programme et contenu :

1. Plan d'expérience
2. Questionnaires
3. Web-scrapping
4. Qualité statistique des données
5. Qualité du format des données

Méthode et organisation pédagogique :

Les enseignements sont prévus pour 46 élèves, les séances de TD ou de TP sont réalisées sur ordinateurs personnels. Le découpage est prévu comme suit :

- 6h de cours (1h par chapitre +1h d'introduction du module)
- 5h de TP (1h par chapitre par chapitre)

Acquis d'apprentissage visés :

Mettre en application un plan d'expérience adapté à la problématique et au contexte pour collecter des données auprès d'utilisateurs ou sur le Web. Vérifier et améliorer la qualité de ces données.

Évaluation : Types d'épreuves et répartition des coef :

- Un contrôle des connaissances acquises sera effectué via un QCM (coef 1).
- Les notions vues dans ce cours seront mises en application dans le projet global du module dont l'évaluation se fera au travers de rapport (coef 2) et de soutenances (coef 1) visant à éclaircir certains points du rapport
- Des évaluations non prévues à l'emploi du temps (contrôles surprise) peuvent advenir.

Retour sur l'évaluation fait à l'élève :

3 semaines après la soutenance :

- Mise à disposition de la correction du QCM
- Retour des rapports et soutenances avec commentaires personnalisés

Support pédagogique et références :

Slides et références

Matière 2 :

Titre de la matière : Bases de données avancées	
Code : 2IA-8.1.2	Titre du module : « Collecte et Stockage de Données »
Semestre : S8	Cursus de rattachement : département 2IA

Heures présentiel	Heures total	Cours	TD	TP	Projet	Contrôles	Travail personnel	Coef /module	ECTS
32	62	12		13	6.25	0.75	30	1.5	2

Titre	Bases de données avancées
résumé	En s'appuyant sur les acquis des étudiants concernant les bases de données, ce cours va plus loin en approfondissant les notions relatives aux bases de données relationnelles, en abordant des notions plus spécifiques : par exemple la gestion de données de grande taille (big data), distribuées (Hadoop), ou les bases de données non structurées (NoSQL), de même que les modes de stockage dématérialisés (<i>cloud</i>). Un projet de mise en application permet de s'assurer que les étudiants savent mettre en pratique les enseignements reçus dans l'intégralité du module.

Responsable	Laurent Bayart – DFA/IMT Mines Alès
Équipe enseignante	Laurent Bayart – DFA/IMT Mines Alès Sébastien Haripse – CERIS/IMT Mines Alès Andon Tchechmedjiev – CERIS/IMT Mines Alès

Mots-clés	Bases de données relationnelles, Bases de données distribuées (Hadoop), cloud, Bases de données non-structurées (NoSQL), Bases de données dématérialisées (<i>cloud</i>).
Prérequis	Introduction aux Bases de données relationnelles – TC

Contexte et objectif général :

Les enseignements des options IASD (Intelligence Artificielle et Science des Données) et IL (Ingénierie Logicielle) reposent sur des données, potentiellement très volumineuses et pas forcément structurées. Cet enseignement présente les principales méthodes et technologie permettant de construire et d'interagir avec différents types de base de données.

Programme et contenu :

1. Base de données relationnelle (*aspects transactionnels, optimisation de requêtes, administration...*)
2. Base de données NoSQL
3. Base de données distribuées (*Hadoop, node4J*)
4. Base de données sur *cloud*

Méthode et organisation pédagogique :

Les enseignements sont prévus pour 46 élèves répartis en deux groupes de 23 élèves pour les séances de TP réalisés sur les ordinateurs personnels de ces derniers.

Le découpage est prévu comme suit :

1. Bases de données relationnelles : 5h de cours + 6 h de TP
2. Bases de données non-structurées : 3h de cours + 3h de TP
3. Bases de données distribuées : 3h de cours + 3h de TP
4. Bases de données dématérialisées : 1h de cours + 1h de TP

Acquis d'apprentissage visés :

Savoir mettre en place et gérer une Base de données adaptée au type de données visées par la problématique.

Évaluation : Types d'épreuves et répartition des coefficients :

- Un contrôle des connaissances acquises sera effectué via un QCM (coef 1)
- Les notions vues dans ce cours seront mises en application dans le projet global du module dont l'évaluation se fera au travers de rapport (coef 2) et de soutenances (coef 1) visant à éclaircir certains points du rapport

- Des évaluations non prévues à l'emploi du temps (contrôles surprise) peuvent advenir.

Retour sur l'évaluation fait à l'élève :

3 semaines après la soutenance :

- Mise à disposition de la correction du QCM
- Retour des rapports et soutenances avec commentaires personnalisés

Support pédagogique et références :

Slides et références

Méthode et organisation pédagogique:

Il s'agit d'un enseignement relativement classique avec une partie réalisée en cours magistral et une partie appliquée au travers de TP et Projets.

Modalité d'évaluation

Le niveau d'acquisition des compétences sera évalué selon les exigences suivantes :

N° indicateur	Indicateur
1	connaître les savoirs formels et pratiques du socle des fondamentaux
2	Exploiter les savoirs théoriques et pratiques
3	Analyser, interpréter, modéliser, émettre des hypothèses, et résoudre

Répartition

Matière	Contrôle	Coefficients	Type de notation	Indicateurs évalués	Chapitres
Plans d'expérience et collecte des données + Bases de données avancées	QCM	1	Individuelle	Tous	Tous
Plans d'expérience et collecte des données + Bases de données avancées	Projet (rapport + soutenance)	3	En binôme (projet)	Tous	Tous

Engagement de l'étudiant, éthique et professionnalisme

La démarche éthique est définie dans le règlement intérieur de l'établissement. Chaque étudiant s'engage à en prendre connaissance et à la respecter.

Obligation des cours : présence obligatoire pour tous à chaque séance.

Nombre d'heures estimées de travail personnel : pour acquérir les compétences demandées, il est nécessaire que l'étudiant consacre minimum 45 min de travail personnel de compréhension et d'approfondissement par séance de cours.
30h de travail personnel pour le projet.

Nombre d'heures estimées de préparation aux travaux dirigés (TD) : 45 min

Pénalité pour retard : Tout travail remis en retard sans motif valable peut être pénalisé de 1 point par jour de retard.

Équipe enseignante

Nom	Domaine d'expertise	Courriel/Téléphone
Nicolas SUTTON-CHARANI	Science de la Donnée, Apprentissage Automatique et Théories de l'Incertain	nicolas.sutton-charani@mines-ales.fr 04 34 24 62 67
Sébastien HARISPE	Apprentissage Automatique, IA et Représentation des Connaissances	sebastien.haripse@mines-ales.fr 04 34 24 62 82
Laurent BAYART	Bases de données relationnelles avancées	laurent.bayart@mines-ales.fr 04 34 24 62 09
Andon TCHECHMEDJIEV	Apprentissage automatique, IA et développement informatique	andon.tchechmedjiev@mines-ales.fr 04 34 24 62 16

ACADEMIC TEACHING	Teaching hours	Coefficients	Crédits
Data Collection and Storage	50 h		
○ Data collection and Design Of Experiment	18	1.5	4
○ Advanced Databases	32	1.5	

Class 1

Class title : Data collection and Design of Experiment	
Code : 2IA-8.1.1	Module title : « Data Collection and Storage »
Semester: S8	Classification : CSAI department

Hours of presence	Total hours	Lectures	Work shop	Labs	Project	Testing	Personal work	Coef /module	ECTS
18	38	6	0	5	6.25	0.75	20	1.5	2

Title	Data collection and Design Of Experiment
Summary	This course aims at laying data collection basis. From the design of experiment to the physical (questionnaires) or virtual (Web) data collection. A special focus is given on data quality (imprecise, missing, outliers). An implementation project proposed further in the module will ensure that students are able to put into practice the knowledge provided in this course.

Head	Nicolas Sutton Charani – CERIS/IMT Mines Alès
Teaching team	Nicolas Sutton Charani – CERIS/IMT Mines Alès Andon Tchechmedjiev – CERIS/IMT Mines Alès

Key words	Design Of Experiment, data collection, web scraping, data quality, outliers
Prerequisites	Linear regression, Python. Students will be brought to use their personal laptops and therefore have to bring them to each session.

Context and general objective:

The courses of DSAI (Data Science and Artificial Intelligence) and SE (Software Engineering) options handle data potentially stored in a structured way. The data processing quality depends on the data quality. This course offers the good practice to achieve this quality.

Programme and contents:

1. Design Of Experiment (DOE)
2. Questionnaires
3. Web-scraping
4. Statistical data quality
5. Computer (format) data quality

Method and pedagogic organisation:

The teachings are planned for 46 students divided into 2 groups for the workshops or Lab session performed on the students' personal laptops.

The organisation is planned as follows:

- 6h of teaching (1h per chapter + 1h of introduction)
- 5h of Labs session (1h per chapter for chapters 1, 2, 3 and 4)

Targeted skills or knowledge :

Implement a design of experiment adapted to a given problem and to its context in order to collect data from users or from the web. Check and improve the resulting data's quality.

Evaluation : Testing procedures used and associated coefficient system :

- multiple-choice questionnaire (coef 1)
- Project: report (coef 2) and defense (coef 1)
- Unscheduled assessments (surprise checks) may occur.

Feedback made to the student :

3 weeks after the defence: i) making MCQ corrections available and ii) reports and defence feedbacks with personalised comments

Teaching material and references :

Slides and references

Class 2
Class title : Advanced Databases
Code : 2IA-8.1.2
Module title : « Data Collection and Storage »
Semester: S8
Classification : CSAI department

Hours of presence	Total hours	Lectures	Work shop	Labs	Project	Testing	Personal work	Coef /module	ECTS
32	62	12		13	6.25	0.75	30	1.5	2

Title	Advanced Databases
Summary	Relying on database basics this course deepens some concepts related to relational databases and tackle more specific problems such as <i>huge, distributed (Hadoop) or non-structured (no-SQL)</i> databases, as well as <i>cloud</i> databases. An implementation project ensures that students are able to put into practice the knowledge provided in this course.

Head	Laurent Bayart – DFA/IMT Mines Alès
Teaching team	Laurent Bayart – DFA/IMT Mines Alès Andon Tchechmedjiev – CERIS/IMT Mines Alès Sébastien Haripse – CERIS/IMT Mines Alès

Key words	Relational databases, distributed databases (Hadoop), non-structured databases (NoSQL), cloud databases.
Prerequisites	Relational databases fundamentals.

Context and general objective:

The courses of DSAI (Data Science and Artificial Intelligence) and SE (Software Engineering) options handle data potentially stored in an unstructured way. The data processing quality depends on the data quality. This course presents the main methods and technologies that enable databases creation and management.

Programme and contents:

1. Relational databases (transactions, requests optimisation, administration, etc.)
2. Non-structured databases (No-SQL)
3. Distributed databases (Hadoop, node4J)
4. Cloud databases

Method and pedagogic organisation:

The teachings are planned for 46 students divided into 2 groups for the Lab sessions performed on the students' personal laptops.

The organisation is planned as follows:

1. Relational databases : 5h of courses + 6h of lab sessions
2. Non-structured databases : 3h of courses + 3h of lab sessions
3. Distributed databases : 3h of courses + 3h of lab sessions
4. Cloud databases : 1h of courses + 1h of lab sessions

Targeted skills or knowledge :

Create and manage a database adapted to the type of data related to a given problem.

Evaluation : The testing procedures used and the associated coefficient system :

- multiple-choice questionnaire (coef 1)
- Project: report (coef 2) and defence (coef 1)
- Unscheduled assessments (surprise checks) may occur.

Feedback made to the student :

3 weeks after the defence:

- making MCQ corrections available
- reports and defence feedbacks with personalised comments

Teaching material and references : Lecture slides and reference textbooks

Method and teaching organisation:

This is a classical course containing a theoretical part with standard courses and a practical one through lab sessions and a project.

Testing procedures

The student's level of knowledge acquisition will be evaluated according to the following points:

N° Indicator	Indicator
1	To know the formal and practical knowledge constituting the foundation of a given field
2	Exploit theoretical and practical knowledge
3	Analyse, interpret, model, hypothesize and solve problems

Grading scheme:

Class	Exam	Coefficients	Administration mode	Evaluated Indicators	Chapters
Data collection and Design of Experiment + Advanced databases	multiple-choice questionnaire	1	Individual (multiple-choice questionnaire)	all	all
Data collection and Design of Experiment + Advanced databases	Project evaluation (report + defense)	3	Individual (multiple-choice questionnaire) and in groups (project)	all	all

Student commitments, ethics and professionalism

Expectations concerning ethics are defined in the establishment's code of conduct. Each student is expected to know and respect the code of conduct.

Obligatory presence in classes: Students must attend all courses, seminars and labs.

Estimated hours of personal study: in order to acquire the required learning level, the student is expected (must) to spend a minimum of 45min of personal study time per hour spent in class.
30h of personal work for the project

Estimated hours of preparation required for labs/Work Shop: 45 min

Late penalties: Late works are subject to penalties as follows: 1 point per day (ratings between 0 and 20).

Teaching team

Name	Field of expertise	Email/phone
Nicolas SUTTON-CHARANI	Data Science, Machine Learning and soft computing	nicolas.sutton-charani@mines-ales.fr 04 34 24 62 67
Sébastien HARISPE	Machine Learning, AI and Knowledge Representation	sebastien.haripse@mines-ales.fr 04 34 24 62 82
Laurent BAYART	Advanced relational databases	laurent.bayart@mines-ales.fr 04 34 24 62 09
Andon TCHECHMEDJIEV	Machine Learning, AI and programming	andon.tchechmedjiev@mines-ales.fr 04 34 24 62 16

Approbation

Ce guide pédagogique entré en vigueur à compter du 07 janvier 2019 a été mis à jour en juillet 2022.
Il est porté à la connaissance des élèves par une publication sur le site de l'école

Rédaction	Vérification	Validation
L'enseignant responsable du module : Nicolas Sutton-Charani	Le responsable d'UE / de département : Sylvie Ranwez	Le directeur de l'école, Pour le directeur et par délégation, Le directeur de la DFA / de la DE : Michel Ferlut