

Partager



Email destinataire :

Vous devez renseigner un email valide

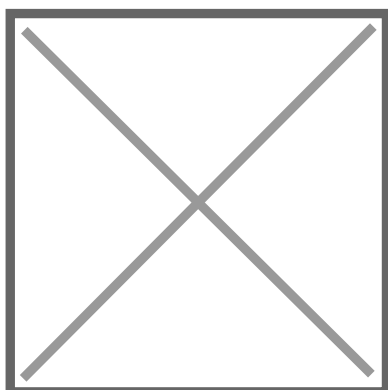
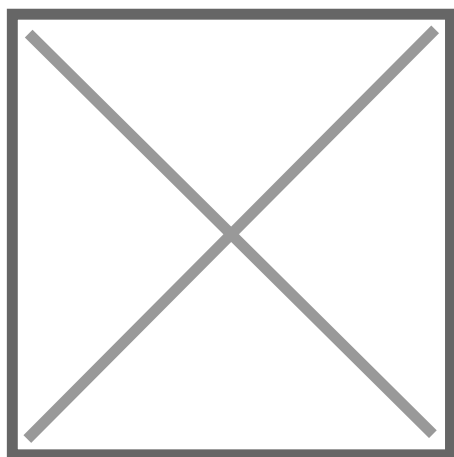
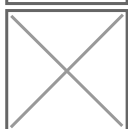
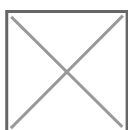
Envoyer

Une erreur est survenue. Merci de réessayer ultérieurement

Le mail de partage a bien été envoyé.

Ajouter au comparateur

M2 DataScale : Gestion de données et extraction de connaissances à large échelle



Master

Mention Informatique

Langue(s) d'enseignement

Français

La formation **DataScale** forme des experts capables de maîtriser toute la chaîne de la donnée : ingénierie des données, architectures, analyse avancée et IA. Elle ouvre la voie à des carrières variées (data scientist, data engineer, IA analyst etc.) dans des secteurs clés et permet également de poursuivre une formation doctorale en recherche publique ou privée.

La formation est dispensée en formation initiale. Les cours ont lieu de mi-septembre à fin mars. Le stage de 5 à 6 mois débute en avril.

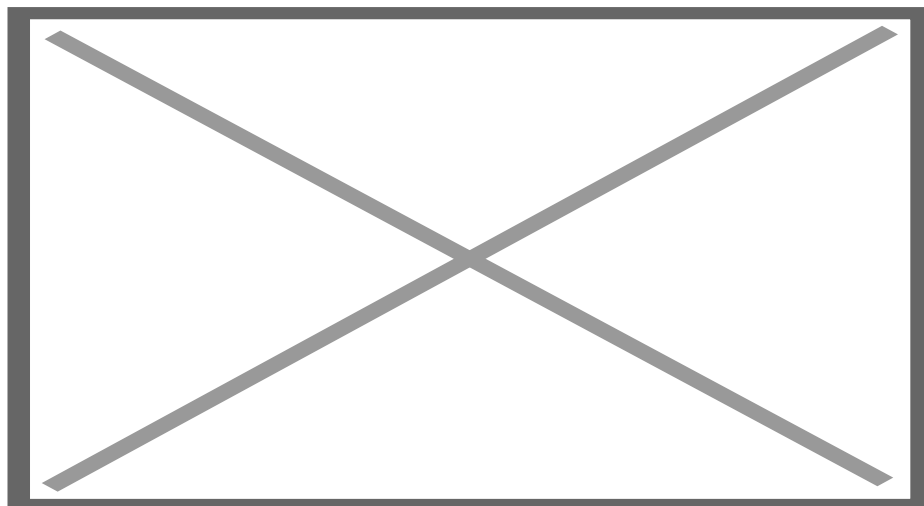
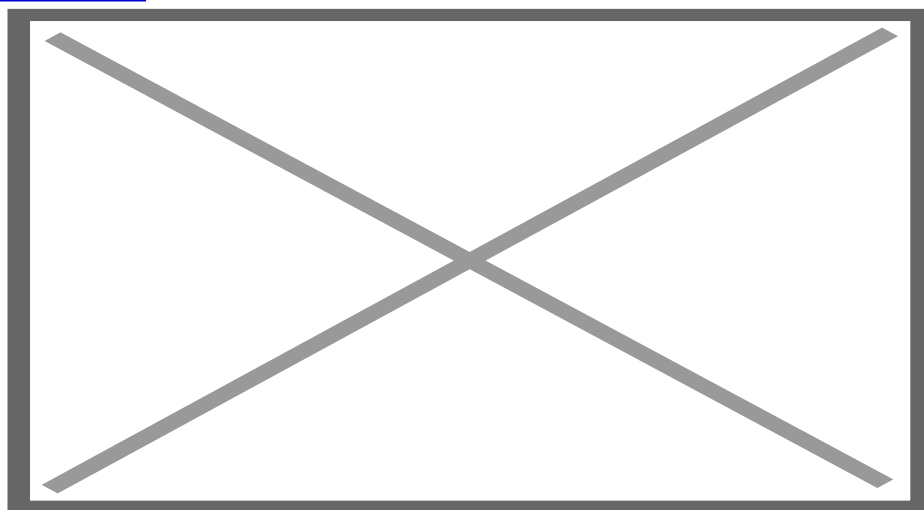
La **‘data’** est au cœur de tous les défis technologiques et de tous les métiers. La formation DataScale vise à offrir aux étudiants une double compétence entre :

- **Ingénierie des données** : conception, administration, sécurisation et optimisation d’architectures multi-échelles (data centers, objets connectés).
- **Analyse des données** : intégration et qualité de données hétérogènes, fouille de données, machine learning et IA.

La formation offre des débouchés multiples (data scientist, data engineer, machine learning engineer ou IA analyst). Les diplômés peuvent poursuivre une formation doctorale dans un laboratoire de recherche, ou intégrer des entreprises de services, des startups, des grands groupes ou des structures de recherche et développement.

[Télécharger la fiche détaillée](#)

[Je candidate](#)





■ ■

Informations

- [Présentation](#)
- [Admission](#)
- [Programme](#)

Présentation

Compétences

- Déployer, utiliser et administrer une infrastructure de gestion de données à large échelle.
- Concevoir et développer des chaînes de traitement permettant de transformer des données brutes en connaissances.
- Extraire, analyser et exploiter l'information et la connaissance stockées dans une infrastructure de gestion de données massives (big data).
- Maîtriser et déployer les algorithmes et techniques de machine learning et d'IA.
- Développer et déployer des applications de gestion de données orientées services.
- Produire et présenter un état de l'art et conduire une démarche scientifique.

Objectifs pédagogiques de la formation

Le parcours DataScale vise à offrir aux étudiants une double compétence très recherchée entre l'ingénierie et l'analyse des données. Les principales thématiques de la formation sont :

- Administration des nouveaux gisements de données, analyse de données de capteurs (domotique, énergie, santé, mobilité), protection de la vie privée, prédiction de l'évolution du climat ou d'une pandémie etc.
- Ingénierie des données : conception, administration, sécurisation et optimisation d'architectures multi-échelles (des data centers aux objets connectés).
- Analyse des données : intégration de données hétérogènes, qualité des données, fouille de données et apprentissage automatique et IA.

La formation offre des débouchés multiples autour de la 'data' dans le monde industriel, les services, les startups, la recherche académique et les laboratoires publics ou privés.

Débouchés :

- Data engineer
- IA analyst
- Data analyst
- Data scientist
- Chief Data Officer
- Administrateur de bases de données (DBA)
- Administrateur de sécurité (DSA)
- Urbaniste de systèmes d'informations
- Architecte / Concepteur / développeur d'applications Big Data, Web, IoT

Agenda

Début de la formation (date indicative)

14/09/2026

Fin de la formation (date indicative)

31/08/2027

Admission

Capacité d'accueil

Places

40

Public visé et prérequis

Le parcours DataScale est proposé en formation initiale uniquement. Les pré-requis correspondent à un niveau équivalent à celui d'un master 1 en informatique avec de solides notions en gestion de bases de données. Les profils types des étudiants qui intègrent le programme DataScale sont les suivants : étudiants ayant obtenu un master en informatique en France, étudiants issus d'écoles d'ingénieurs partenaires en dernière année inscrits dans un double cursus, étudiants titulaires d'un master en ingénierie ou en informatique obtenu à l'étranger.

Les profils typiques d'étudiants intégrant la formation DataScale sont : étudiants ayant obtenu un master 1 informatique en France, étudiants d'écoles d'ingénieurs partenaires effectuant leur dernière année en double cursus, étudiants ayant déjà obtenu un diplôme d'ingénieur ou de master informatique à l'étranger.

Période(s) de candidature

Plateforme Inception

Du 16/02/2026 au 10/07/2026

Pièces justificatives

Obligatoires

Copie diplômes.

Lettre de motivation.

Tous les relevés de notes des années/semestres validés depuis le BAC à la date de la candidature.

Attestation de français (obligatoire pour les non francophones).

Curriculum Vitae.

Facultatives

Document justificatif des candidats exilés ayant un statut de réfugié, protection subsidiaire ou protection temporaire en France ou à l'étranger (facultatif mais recommandé, un seul document à fournir) :

- Carte de séjour mention réfugié du pays du premier asile
- OU récépissé mention réfugié du pays du premier asile
- OU document du Haut Commissariat des Nations unies pour les réfugiés reconnaissant le statut de réfugié
- OU récépissé mention réfugié délivré en France
- OU carte de séjour avec mention réfugié délivré en France
- OU document faisant état du statut de bénéficiaire de la protection subsidiaire en France ou à l'étranger.

Programme

Tronc Commun

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet Tutorat
Machine Learning	2.5	Semestre 1	21							
■										

Machine Learning

ECTS : 2.5

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 21

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Notions de statistiques et d'algèbre linéaire.

Nature de l'évaluation

Evaluation Continue non Intégrale

[Qualité des données](#)

2.5 Semestre 1 21

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
■											

Qualité des données

ECTS : 2.5

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 21

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Concepts fondamentaux de la gestion de données.

Nature de l'évaluation

Evaluation Continue non Intégrale

[Modèles et éco-
systèmes Post-
Relationnels](#)

2.5

Semestre 1

21

0

■

Modèles et éco-systèmes Post-Relationnels

ECTS : 2.5

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 21

Travaux dirigés : 0

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Maîtrise des concepts des bases de données relationnelles et des systèmes de gestion de bases de données, maîtrise d'un langage de programmation orienté-objet (Java, Python, ...), maîtrise des outils systèmes (shell, ...) et des outils de développement.

Nature de l'évaluation

Evaluation Continue non Intégrale

[Intégration de
données](#)

2.5

Semestre 1

21

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
■											

Intégration de données

ECTS : 2.5

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 21

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Bases de données : création et manipulation de bases de données relationnelles, mécanismes internes des SGBD relationnels.

Nature de l'évaluation

Evaluation Continue non Intégrale

Architectures orientées Services	2.5	Semestre 1	21	0
--	-----	------------	----	---

■

Architectures orientées Services

ECTS : 2.5

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 21

Travaux dirigés : 0

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Bonne connaissance des concepts de base de la programmation orientée-objet, et du développement d'applications Web.

Nature de l'évaluation

Evaluation Continue non Intégrale

Architectures des gestionnaires de données	2.5	Semestre 1	21	0	0
--	-----	------------	----	---	---

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
■											

Architectures des gestionnaires de données

ECTS : 2.5

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 21

Travaux pratiques : 0

Travaux dirigés : 0

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Niveau M1 informatique ou 4ème année d'école d'ingénieur ayant reçu une formation en :

- bases de données : concepts fondamentaux, SGBD relationnels, principes de base de l'indexation et de l'évaluation de requêtes
- systèmes distribués : concepts fondamentaux, principes de base du contrôle de concurrence et de la réplication

Nature de l'évaluation

Evaluation Continue non Intégrale

Data Analysis

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
Fouille de données	2.5	Semestre 1	21								

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
■											

Fouille de données

ECTS : 2.5

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 21

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Notions de statistiques et d'analyse de données.

Nature de l'évaluation

Evaluation Continue non Intégrale

[Web sémantique](#) 2.5 Semestre 1 21

■

Web sémantique

ECTS : 2.5

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 21

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Maîtrise des concepts de la gestion de données, maîtrise d'un langage de programmation (Java ou Python).

Nature de l'évaluation

Evaluation Continue non Intégrale

[Modélisation de processus métiers](#) 2.5 Semestre 1 21 0

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
■											

Modélisation de processus métiers

ECTS : 2.5

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 21

Travaux dirigés : 0

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Bonne connaissance des concepts de base des systèmes d'information, les architectures SOA, et du développement d'applications Web.

Nature de l'évaluation

Evaluation Continue non Intégrale

[Analyse de masses de données de mobilité](#)

2.5

Semestre 1 21

■

Analyse de masses de données de mobilité

ECTS : 2.5

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 21

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Advance database techniques (query processing, data integration and quality) Machine learning (recommended)

Nature de l'évaluation

Evaluation Continue non Intégrale

Professionnalisation

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
----------	------	----------	-------	----	----	--------------	--------------	-----------	---------------	--------	---------

Séminaires Industriels et Recherche	1	Semestre 2	20
■			

Séminaires Industriels et Recherche

ECTS : 1

Semestre calendaire : Semestre 2

Détail du volume horaire

Travaux dirigés : 20

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Aucun

Nature de l'évaluation

Evaluation Continue non Intégrale

Projet Programmation	1.5	Semestre 2	1.5
--------------------------------------	-----	------------	-----

■

Projet Programmation

ECTS : 1.5

Semestre calendaire : Semestre 2

Détail du volume horaire

Travaux dirigés : 1.5

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

UE 'projet conception'

Nature de l'évaluation

Evaluation Continue non Intégrale

Projet Conception	1.5	Semestre 2	1.5
-----------------------------------	-----	------------	-----



Projet Conception

ECTS : 1.5

Semestre calendaire : Semestre 2

Détail du volume horaire

Travaux dirigés : 1.5

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Suivi des UE de tronc commun DataScale

Nature de l'évaluation

Evaluation Continue non Intégrale

Anglais	3	Semestre	27
		2	



Anglais

ECTS : 3

Semestre calendaire : Semestre 2

Détail du volume horaire

Travaux dirigés : 27

Langue d'enseignement

Français

Enseignement à distance

non

Nature de l'évaluation

Evaluation Continue non Intégrale

Régulations des données et des contenus numériques	1	Semestre	9
		2	



Sécurité des données corporate et personnelles

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
■											

Sécurité des données corporate et personnelles

ECTS : 2.5

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 21

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Bonnes connaissances des SGBD

Concepts de base de la cryptographie (un rappel sera effectué)

Nature de l'évaluation

Evaluation Continue non Intégrale

[Fondements des](#)

[langages](#)

[d'interrogation de](#)

[données](#)

2.5

Semestre 1

21

Fondements des langages d'interrogation de données

ECTS : 2.5

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 21

Langue d'enseignement

Anglais

Enseignement à distance

non

Prérequis

Notions de logique

Objectifs d'apprentissage

The course aims to provide a principled introduction to query processing, its logical foundations, and the recent query language evolution. It will cover the basics of graph processing: relational querying, Datalog for expanding its abilities to handle graphs, and Cypher, the most commonly currently used graph query language.

Organisation générale et modalités pédagogiques

It introduces key notions from first-order logic and gives a gentle overview of the fundamentals of the relational data model, with a focus on relational algebra and calculus. It then delves deeper into an analysis of the logical foundations of query languages. As such, it provides a detailed presentation of Datalog [6], which is a recursive query language that extends relational algebra with recursion, and that has been used to express a wide spectrum of modern data management tasks. We cover the standard Datalog language, its extension with negation, as well as application-specific dialects (with aggregates, custom predicates, and internalized transitive closure). Focusing on some of these fragments, we survey their various semantics [1, 8, 2]. We then discuss corresponding evaluation techniques and optimizations [5]. In the second part of the course, we concentrate on the connections between Datalog and graph query languages [7, 2]. We illustrate these through practical, interactive hands-on examples centered around Neo4j's Cypher query language. We will unravel basic and advanced features of the language through a real-world application scenario. Finally, we will outline current research advances in designing constraint and schema languages for graph databases [4, 3].

Bibliographie

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
Gestion de données et de services dans le cloud	2.5	Semestre 1	21	0							
■											

Gestion de données et de services dans le cloud

ECTS : 2.5

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 21

Travaux dirigés : 0

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Bonne connaissance des architectures à bases de services, des concepts fondamentaux des bases de données, des applications client/serveur.

Nature de l'évaluation

Evaluation Continue non Intégrale

Indexation et visualisation de données massives	2.5	Semestre 1	21
---	-----	------------	----

Indexation et visualisation de données massives

ECTS : 2.5

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 21

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Bonne connaissance des SGBD et des éco-systèmes post-relationnels.

Objectifs d'apprentissage

La formation présente un état de l'art des concepts d'indexation et de visualisation de données massives. Elle intègre également la mise en oeuvre de la pile Elastic Stack (Elasticsearch, Logstash, Kibana) afin de permettre aux experts métiers de questionner les différents fichiers de traces produits par un système d'information. Il est alors possible de gérer, d'exploiter et de visualiser les logs de tous type d'application et d'OS.

Organisation générale et modalités pédagogiques

- Connaître les principes fondamentaux de l'indexation et de la visualisation de données massives
- Comprendre la pile ELK (Elasticsearch, logstash, Kibana) et quand l'utiliser
- Installer et configurer Elasticsearch, logstash et Kibana
- Savoir utiliser Elasticsearch, logstash et Kibana pour indexer, chercher et visualiser des données et des documents (projet)
- Produire des tableaux de bord
- Intégrer la pile avec Hadoop et Spark et Kafka

Nature de l'évaluation

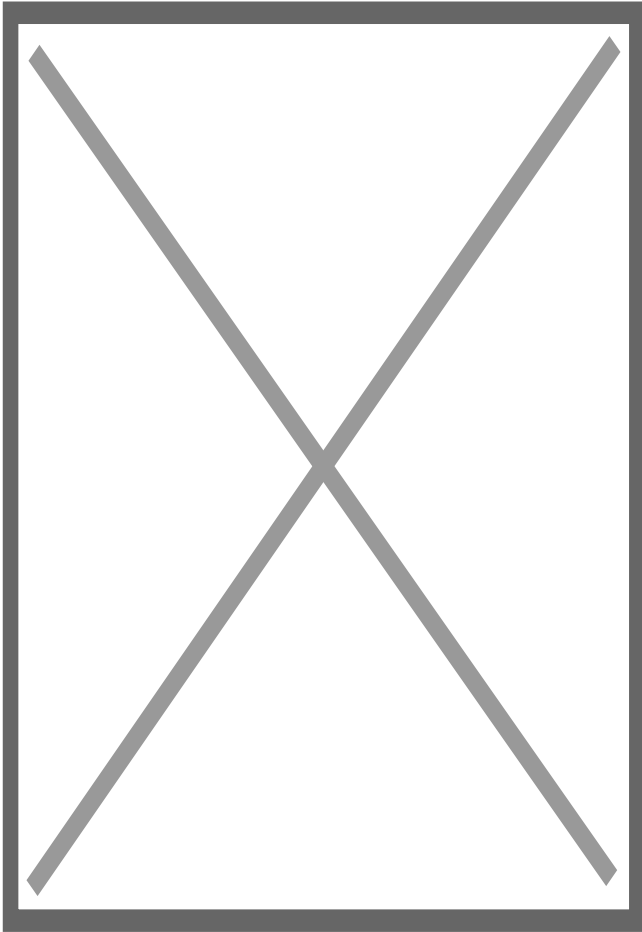
Evaluation Continue non Intégrale

Lieu(x) d'enseignement

VERSAILLES

Campus de la formation

Javascript est requis pour visualiser cette carte.



Versailles

Versailles

Bus 471, 5371, 6179, 6201, 6203, 6204, 6205, 6217

Versailles Rive Droite - Ligne L

RER C Versailles Château Rive gauche

[Voir](#)

Restaurant CROUS

Bibliothèque

Résidence étudiante

Contact(s)

- **Zoubida KEDAD**

PR
- **Mustapha LEBBAH**

PR
- **Clémence Perez**

Secrétaire Département d'informatique

<http://www.universite-paris-saclay.fr/formation/master/informatique/m2-datascale-gestion-de-donnees-et->

