



**Data ScienceTech Institute**

---

## **Formation continue**

Éligible aux mécanismes de financement de la formation professionnelle en France (CPF, Transitions Pro, Dispositifs d'Aides Régionales, etc.)

# TABLE DES MATIÈRES

---

<b>01</b>	Formation Continue	p3
<b>02</b>	Processus	p5
<b>03</b>	Comment créer son module ?	p6
<b>04</b>	Applied MSc in Data Analytics	p7
<b>05</b>	Applied MSc in Data Engineering for Artificial Intelligence	p8
<b>06</b>	Applied MSc in Data Science & Artificial Intelligence	p9
<b>07</b>	Les différents modules d'exemples	p10
<b>08</b>	Frais de formation	p11



# FORMATION CONTINUE



Admission mensuelle



En ligne



À votre rythme

Étudiez la **Data Analytics**, **Data Science** ou le **Data Engineering** en choisissant l'un de nos cours modulaires (certifiés au RNCP de niveau 7). Ces cours peuvent être suivis en ligne et à votre propre rythme.

Nos modules de formation vous permettront de commencer à tout moment et d'acquérir immédiatement de nouvelles compétences. Comme le mode asynchrone (SPOC), ce parcours d'études est destiné aux professionnels qui souhaitent travailler et étudier en même temps.

## Description

Les programmes « Applied MSc » de Data ScienceTech Institute proposent exclusivement une validation totale des connaissances et compétences associées à nos diplômes et titre RNCP 34262 « Expert en Sciences des Données ».

Notre proposition de « **formation continue** » répond à un souhait de **validation partielle** mais aussi **capitalisable** depuis ces programmes.

Les cours proposés, sous forme modulaire, permettent :

- de rentrer dans le même processus valorisant d'admission sélective.
- de suivre les mêmes contenus d'enseignement.
  - en asynchrone, dans le mode « Self-Paced Online Course » (SPOC), pour les personnes ne pouvant pas libérer du temps professionnel.
  - en synchrone, selon les calendriers de cours DSTI.
  - ou un mix des deux.
- de se soumettre aux mêmes évaluations.
- d'obtenir un relevé de notes pour les cours validés, capitalisable par crédits European Credits Transfer System (ECTS) envers un programme complet « Applied MSc » et son titre RNCP niveau 7 (master) associé.
- éligibles au Compte Personnel de Formation (CPF), puisque systématiquement associé à un titre RNCP.

La composition de votre parcours de formation à validation partielle est entièrement individualisée, fonction des prérequis de compétences (évalués dans notre processus d'admission), vos besoins d'évolution professionnelle et potentiellement, votre désir d'obtenir un diplôme DSTI et titre RNCP associé par capitalisation de crédits ECTS.

Les éléments ci-dessous présentent les différentes modalités de composition d'un parcours de formation continue en validation partielle de nos programmes Applied MSc.



## Entretien d'accompagnement

S'inscrire à l'un de nos cours modulaires demande réflexion et engagement. Il est important dans vos critères de sélection que le cursus soit réellement en adéquation avec vos besoins et vos projets professionnels. **Ainsi, nous offrons un soutien pédagogique personnalisé avant même le début du programme.**

Avant le début du module, un entretien avec un membre de la Direction des Études a lieu afin de :

Vérifier que le candidat possède les prérequis nécessaires pour suivre le module. Si nécessaire, le candidat se verra proposer, sans frais supplémentaires, des vidéos ou du matériel supplémentaire pour se mettre à niveau sur les compétences requises.

Confirmer avec le candidat l'adéquation du module avec ses objectifs professionnels. Si nécessaire, il sera possible d'ajuster le contenu du module avec d'autres cours pour répondre aux besoins du candidat.

## Des parcours d'apprentissage personnalisés

Les formations proposées par DSTI s'adressent aux cadres, demandeurs d'emplois ou salariés désireux de maîtriser les compétences et les outils nécessaires au développement de leur carrière.

Ils pourront ainsi profiter des nombreuses opportunités offertes par les domaines de la science des données et du big data.

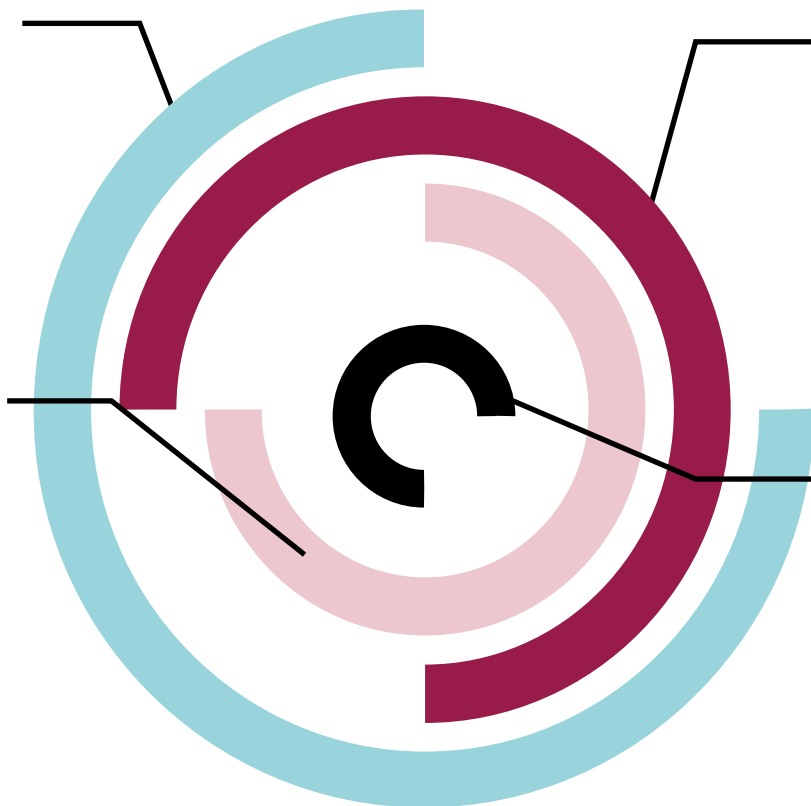
# PROCESSUS

## 1) AVANT L'ENTRETIEN :

Approfondir son projet de formation lors d'un entretien personnalisé.  
Vérifier que les besoins et attentes sont en adéquation avec le module envisagé.  
Invitation à passer notre examen d'admissibilité.

## 3) PARCOURS PERSONNALISÉ :

Une fois que les prérequis ont été validés, il ne reste plus qu'à confirmer avec le candidat le module choisi. Il a la possibilité de choisir différentes volumétries et différents cours dans les 3 formations.



## 2) PENDANT L'ENTRETIEN :

Vérifiez que le candidat possède les prérequis nécessaires pour suivre le module.  
Si nécessaire, le candidat se verra proposer, sans frais supplémentaires, des vidéos ou du matériel supplémentaire pour se mettre à niveau sur les compétences requises.

## 4) APRÈS L'ENTRETIEN :

Il ne reste plus qu'à acquérir les nouvelles compétences et booster votre carrière professionnelle.

## OBJECTIFS :

---

- Maîtriser les compétences et les outils nécessaires au développement de sa carrière.
- Profiter des nombreuses opportunités offertes par les domaines de la science des données et du big data.

## POINTS FORTS :

---

- Un accompagnement personnalisé tout au long de votre formation à DSTI.
- Un suivi de direction des études propre à l'enseignement supérieur.

# COMMENT CRÉER SON MODULE ?

## Applied MSc in Data Analytics

- Applied Mathematics for Data Science (25 hrs)
- CRM Data Management (25 hrs)
- Big Data Processing with R (25 hrs)

Voir l'ensemble du programme page 6

## Applied MSc in Data Engineering for Artificial Intelligence

- Data Wrangling with SQL (25 hrs)
- The Hadoop Ecosystem (50 hrs)
- Document Databases (NoSQL - Part 2) (10 hrs)
- Data Pipeline Part 1 & 2 (50 hrs)

Voir l'ensemble du programme page 7

## Applied MSc in Data Science & Artificial Intelligence

- Continuous Optimisation (25 hrs)
- Semantic Web technologies (25 hrs)
- Foundations of Statistical Analysis & Machine Learning, Part 2 (40 hrs)
- Survival Analysis using R (25 hrs)

Voir l'ensemble du programme page 8

### Mon module - 75 h

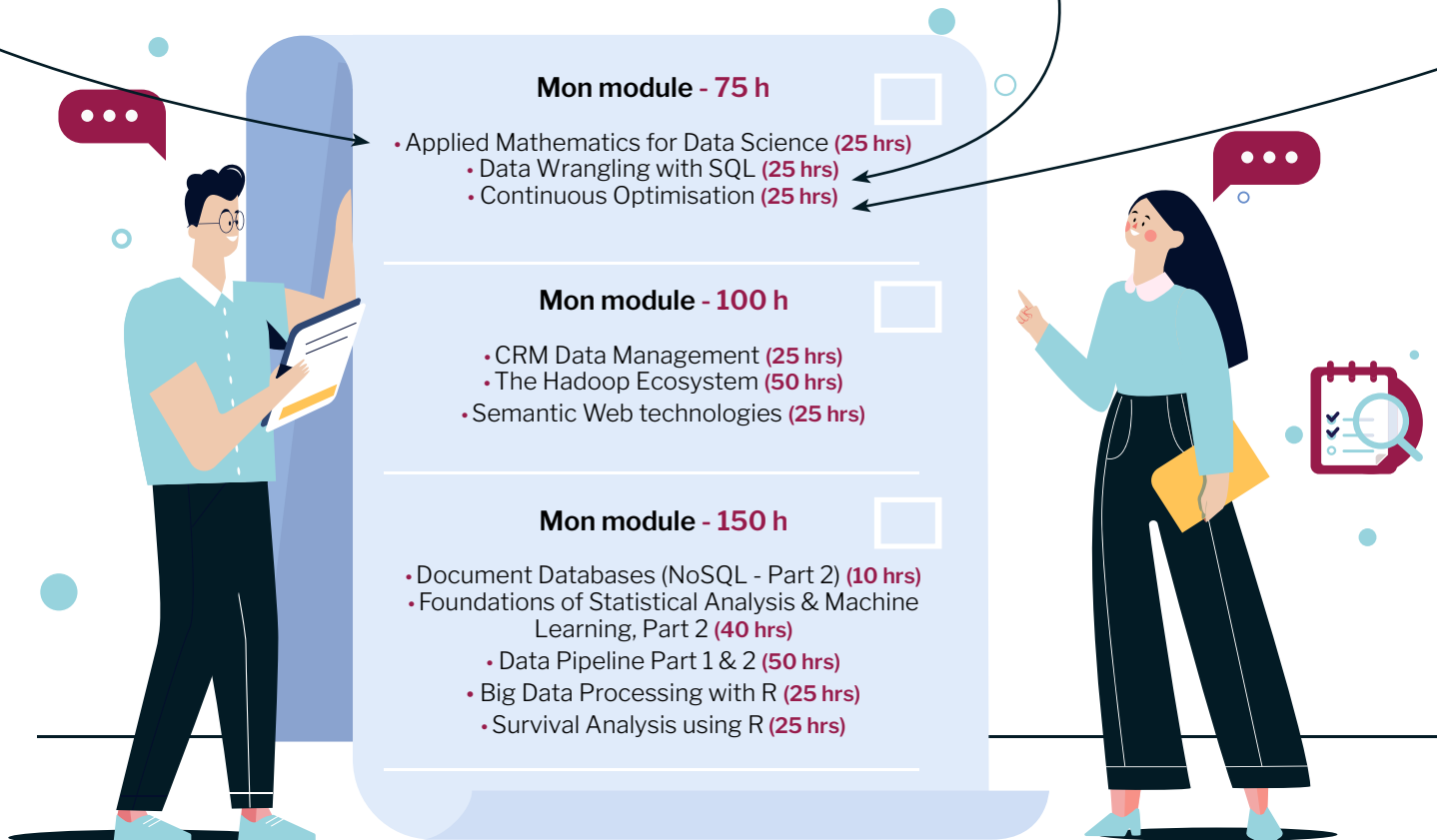
- Applied Mathematics for Data Science (25 hrs)
- Data Wrangling with SQL (25 hrs)
- Continuous Optimisation (25 hrs)

### Mon module - 100 h

- CRM Data Management (25 hrs)
- The Hadoop Ecosystem (50 hrs)
- Semantic Web technologies (25 hrs)

### Mon module - 150 h

- Document Databases (NoSQL - Part 2) (10 hrs)
- Foundations of Statistical Analysis & Machine Learning, Part 2 (40 hrs)
- Data Pipeline Part 1 & 2 (50 hrs)
- Big Data Processing with R (25 hrs)
- Survival Analysis using R (25 hrs)



# PROGRAMME : Applied MSc in Data Analytics

## 1/ Data Analytics - 125 heures / 9 ECTS

### • Applied Mathematics for Data Science (25 hrs)

Calculus - Linear Algebra - Trigonometry & Complex Numbers

### • Foundations of Statistical Analysis & Machine Learning Part I (25 hrs)

Descriptive Statistics - Probability Theory – Applications using R

### • Big Data Processing with R (25 hrs)

Import and manipulate very large datasets with R - Best data structures selection - Data Transformation - Visualisation - exploring and modelling

### • Python Machine Learning Labs (30 hrs)

Data structures - Cleaning & preparation - Pandas - Matplotlib - Scikit-learn - OpenCV - Python & Flask - Keras - Numpy

### • Semantic Web Technologies (25 hrs)

Representing and querying web-rich data (RDF, SPARQL) - Introducing Semantics in Data (RDFS, Ontologies) - Tracing and following data history (VOiD, DCAT, PROV-O)



## 2/ Databases - 110 heures / 9 ECTS

### • Data Wrangling with SQL (25 hrs)

Fundamentals of Relational Model & Databases, Relational Algebra - Advanced SQL queries, stored procedures & triggers (T-SQL), Dynamic SQL - Applications with Microsoft SQL Server

### • Datawarehousing & ETL (25 hrs)

Using Microsoft SQL Server: Stand-alone and cluster deployments, design & implementation of a datawarehouse, structuring an Extract - Transform - Load process...

### • Graph Databases (NoSQL - Part 1) (25 hrs)

Preparation to Neo4j Certification - Modelling a graph-based problem - Implementation with the Neo4j database

### • Document Databases (NoSQL - Part 2) (10 hrs)

Fundamentals of MongoDB Databases - Collection & Document Advanced Mongo DB queries - MongoDB aggregations - Mongo DB data architecture - Applications with MongoDB and Robo3T

### • Data Pipeline Part I (25 hrs)

XML dataflow – DTD & Schemas – XLS Transformation – JSON & Transformations

## 3/ Data Management and Visualisation - 100 heures / 10 ECTS

### • Advanced Excel for Data Analytics & Machine Learning (25 hrs)

Formulas - Data Visualisation - PowerPivot - Solver - Visual Basic for Application

### • Data & Machine Learning Visualisation Ecosystem (25 hrs)

Tables & Machine Learning with SAS Viya

### • Reporting & Visualisation (25 hrs)

Preparation of certification “Analyzing Data with Microsoft BI (DA-100)”

### • CRM Data Management (25 hrs)

Preparation of certification “Microsoft Power Platform Functional Consultant (PL-200)”



## 4/ Management - 50 heures / 2 ECTS

### • Data Laws & Regulations - Philosophies, Geopolitics & Ethics (25 hrs)

EU & USA approaches - GDPR - Safe Harbour & Successors - Common Law vs Code law

### • IT Project Management: PMP-PMI and Agile Approaches & Software Quality (25 hrs)

PMBOK (PMI) - Agile Approaches - Kanban - Quality Metrics - Unit & Integration testing

# PROGRAMME : Applied MSc in Data Engineering for Artificial Intelligence

## 1/ Distributed & Performance IT - 200 heures / 10 ECTS

### • Amazon AWS “Cloud-Computing DSTI Chair” (50 hrs)

Preparation to AWS Certified Solutions Architect – Associate Certification

### • Cloud Computing - Microsoft Azure (25 hrs)

Comparative overview of with Amazon AWS – Focus on Azure Services specific to data lakes and data pipelines

### • Semantic Web technologies (25 hrs)

Representing and querying web-rich data (RDF, SPARQL) – Introducing Semantics in Data (RDFS, Ontologies) – Tracing and following data history (VOiD, DCAT, PROV-O)

### • Software Engineering - Part 1 (25 hrs)

Review of programming and memory management fundamentals - introduction to Microsoft .NET environment – Applications in C & C#

### • Software Engineering - Part 2 (25 hrs)

Fundamentals of algorithmics & data structures using object-oriented programming – Applications in C++ & Python

### • Python Machine Learning Labs (30 hrs)

Data structures – Cleaning & preparation – Pandas – Matplotlib – Scikit-learn – OpenCV – Python & Flask – Keras – Numpy

### • Java & Scala programming (25 hrs)

Java for Map Reduce in Hadoop & Scala for SPARK



## 2/ Data Management - 185 heures / 8 ECTS

### • Data Wrangling with SQL (25 hrs)

Fundamentals of Relational Model & Databases - Relational Algebra – Advanced SQL queries - stored procedures & triggers (T-SQL) - Dynamic SQL – Applications with Microsoft SQL Server

### • Data warehousing & ETL (25 hrs)

Using Microsoft SQL Server: stand-alone and cluster deployments - design and implementation of a datawarehouse - structuring an Extract-Transform-Load process – Applications with Microsoft SQL Server

### • Graph Databases (NoSQL - Part 1) (25 hrs)

Preparation to Neo4j Certification – Modelling a graph-based problem, implementation with the Neo4j database

### • Document Databases (NoSQL - Part 2) (10 hrs)

Fundamentals of MongoDB Databases, Collection and Document – Advanced MongoDB queries, MongoDB aggregations, MongoDB data architecture – Applications with MongoDB and Robo3T

### • The Hadoop Ecosystem (50 hrs)

HDFS – Scheduling & resources management – Workflow management & ETL – Dataflow management – Scalable Enterprise Serial Bus – Realtime processing with SPARK – Machine Learning – Data Exploration & Visualisation

### • Data Pipeline Part 1 & 2 (50 hrs)

XML dataflow - DTD & Schemas - XLS Transformation - JSON & Transformations – Cloud-based solutions with Glue in AWS & AWS Kinesis – Open-source solutions with Apache Kafka & Beam

## 3/ Operational Methodologies - 150 heures / 4-6 ECTS

### • Data Laws & Regulations - Philosophies, Geopolitics & Ethics (25 hrs)

EU & USA approaches – GDPR – Safe Harbour & Successors – Common Law vs Code Law

### • IT Project Management: PMP - PMI and Agile Approaches (25 hrs)

PMBOK (PMI) – Agile Approaches – Kanban – Quality Metrics

### • CRM Data Management \* (25 hrs)

Preparation of certification "Microsoft Power Platform Functional Consultant (PL-200)"

### • DevOps & Continuous Integration (50 hrs)

The DevOps toolbox: Nagios, Consul, Docker, Ansible, GitHub – Continuous Integration with Jenkins & Kubernetes

### • Cybersecurity (25 hrs)

System Security Design Patterns – Infrastructure security – Data at-rest and in-transit encryption – Code safety



## 4/ Data Science - 125 heures / 4-8 ECTS

### • Applied Mathematics for Data Science (25 hrs)

Calculus – Linear Algebra – Trigonometry & Complex Numbers

### • Foundations of Statistical Analysis and Machine Learning - Part 1 (25 hrs)

Descriptive Statistics - Probability Theory – Applications using R

### • Big Data Processing with R (25 hrs)

Import and manipulate very large datasets with R – Best data structures selection – Data Transformation – Visualisation – Exploring and modelling

### • Artificial Neural Networks \* (25 hrs)

Perceptron's layers, weights, biases – Hyperparameter – Activation and cost functions – Review of optimization algorithms – Backpropagation – Learning mechanism – Classification & regression – Applications in Python using TensorFlow

### • Deep Learning (25 hrs)

Recurrent Neural Networks, LSTM, Residual Networks, Computer Vision & NLP – Deep Learning on GPU – Application using Python & PyTorch



# PROGRAMME : Applied MSc in Data Science & Artificial Intelligence

## 1/ Core Data Science & Artificial Intelligence - 190 heures / 8-10 ECTS

### • Applied Mathematics for Data Science (25 hrs)

Calculus – Linear Algebra – Trigonometry & Complex Numbers

### • Foundations of Statistical Analysis & Machine Learning, Part 1 (25 hrs)

Descriptive Statistics - Probability Theory – Applications using R

### • Foundations of Statistical Analysis & Machine Learning, Part 2 (40 hrs)

Tests – Estimators – Confidence Intervals – Inference – ANOVA – PCA – Simple Linear Regression – Applications using R

### • Time-Series Analysis (25 hrs)

Mathematical foundations – Applications with R including using Neural Networks

### • SAS "The SAS Ecosystem DSTI Chair" (25 hrs)

Preparation to SAS BASE Certification - SAS BASE programming, application with SAS STATS

### • Continuous Optimisation (25 hrs)

Critical points – Multiple variables function optimisation – Gradient methods – Constraint-based optimisation with Lagrange Multiplier – Applications Python

### • Artificial Neural Networks (25 hrs)

Perceptron's layers, weights, biases – Hyperparameter – Activation and cost functions – Review of optimization algorithms – Backpropagation – Learning mechanism – Classification & regression – Applications in Python using TensorFlow



## 2/ Core Data Engineering - 200 heures / 9 ECTS

### • Software Engineering Part 1 (25 hrs)

Fundamentals of algorithmics & data structures using classical design & programming – Applications in C

### • Software Engineering Part 2 (25 hrs)

Fundamentals of algorithmics & data structures using object-oriented programming – Applications in C++ & Python

### • Python Machine Learning Labs (30 hrs)

Data structures – Cleaning & preparation – Pandas – Matplotlib – Scikit-learn – OpenCV – Python & Flask – Keras – Numpy

### • Data Wrangling with SQL (25 hrs)

Fundamentals of Relational Model & Databases, Relational Algebra – Advanced SQL queries, stored procedures & triggers (T-SQL), Dynamic SQL – Applications with Microsoft SQL Server

### • Amazon AWS "Cloud-Computing DSTI Chair" (50 hrs)

Preparation to AWS Certified Solutions Architect – Associate Certification

### • The Hadoop & Spark Ecosystem (50 hrs)

HDFS – Scheduling & resources management – Workflow management & ETL – Dataflow management – Scalable Enterprise Serial Bus – Realtime processing – Machine Learning – Data Exploration & Visualisation – Application using a real, industry-grade Hadoop & Spark cluster

## 3/ Applied Data Science & Artificial Intelligence - 210 heures / 9-11 ECTS

### • Advanced Statistical Analysis and Machine Learning (35 hrs)

Multiple Linear Regression - CART and Random Forests and applications to their applications - Features Selection & Engineering - Models Comparison & Competition Applications using R

### • Statistical Analysis of Massive and High Dimensional Data (25 hrs)

Context for new uses of massive datasets (open data, social networks, Twitter...) – Review of conventional statistical methods (tests, regression, classification) and their (un)suitability massive datasets – Latest alternative statistical tools for analysing modern datasets – Implementation in realistic situations using R

### • Survival Analysis using R (25 hrs)

Statistical analysis of duration data - Applications using R

### • Inverse Problems & Data Assimilation \* (25 hrs)

Variational and sequential data assimilation – Identification of the initial condition, parameter estimation – Applications using Python

### • Graph Databases - NoSQL (25 hrs)

Preparation to Neo4j Certification – Modelling a graph-based problem, implementation with the Neo4j database

### • Deep Learning (25 hrs)

Recurrent Neural Networks - LSTM - Residual Networks - Computer Vision & NLP – Deep Learning on GPU – Application using Python & PyTorch

### • Agent-Based Modeling (25 hrs)

Solving complex problems using ABM – Comparisons with statistical – Markov and system dynamics approaches – ABM validation for "trustability"

### • Semantic Web technologies (25 hrs)

Representing and querying web-rich data (RDF, SPARQL) – Introducing Semantics in Data (RDFS, Ontologies) – Tracing and following data history (VOID, DCAT, PROV-O)



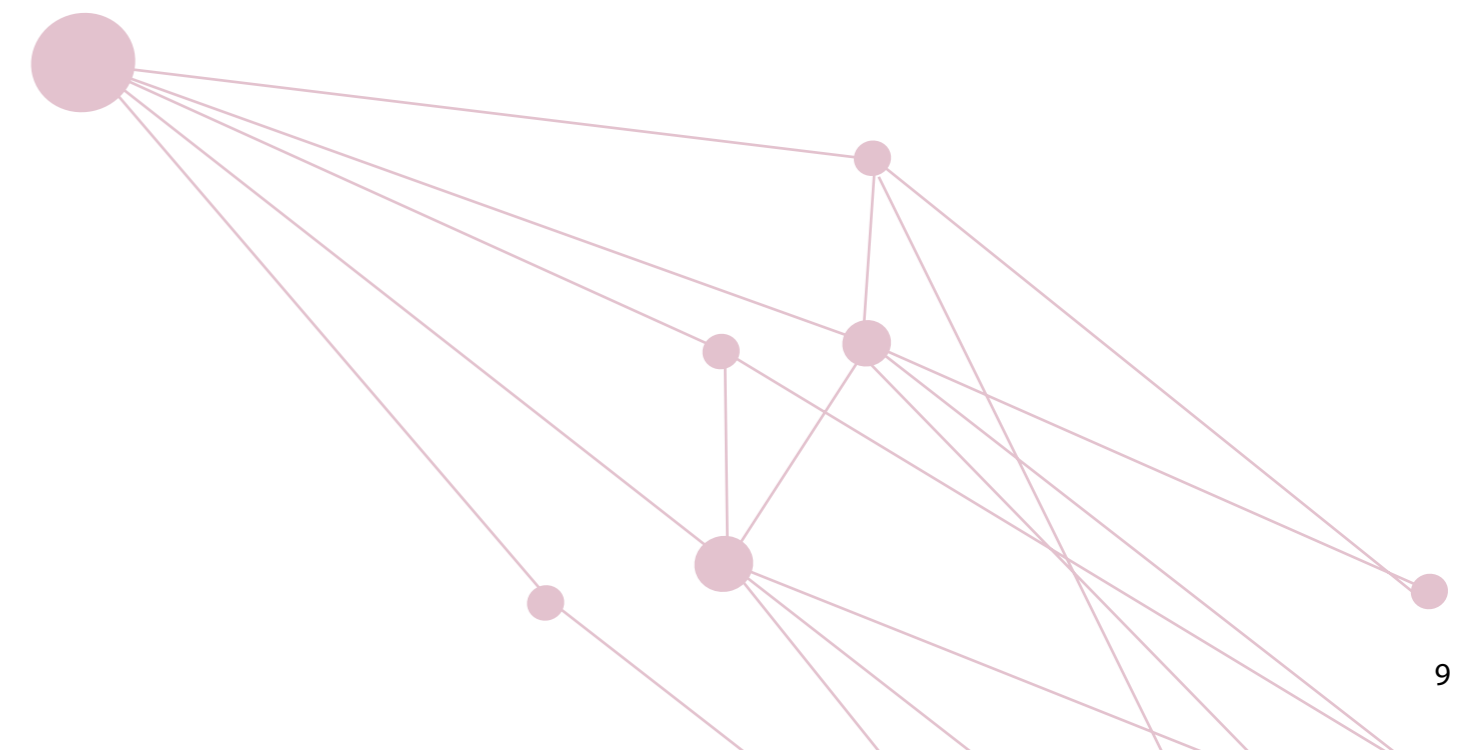
## 4/ Management, Ethics & Laws - 50 heures / 2 ECTS

### • Data Laws & Regulations - Philosophies, Geopolitics & Ethics (25 hrs)

EU & USA approaches – GDPR – Safe Harbour & Successors – Common Law vs Code Law

### • IT Project Management: PMP-PMI and Agile Approaches (25 hrs)

PMBOK (PMI) – Agile Approaches – Kanban – Quality Metrics



# LES DIFFÉRENTS MODULES D'EXEMPLES

## DATA ANALYST

### Data Analytics - *Initiation* (50 hrs)

Python Machine Learning Labs (25 hrs)  
or Advanced Excel for Data Analytics & Machine Learning (25 hrs)  
Data Wrangling with SQL (25 hrs)

### Databases (75 hrs)

Relational databases for ETL & Data Warehousing (25 hrs)  
NoSQL (25 hrs)  
Data Pipeline 1 – XML & JSON (25 hrs)

### Machine Learning Analyst (100 hrs)

Foundations of Statistical Analysis & Machine Learning – Part 1 (25 hrs)  
R for Big Data (25 hrs)  
Data & Machine Learning Visualisation Ecosystem (25 hrs)  
Reporting & Visualisation – Preparation to Microsoft PowerBI Certification (25 hrs)

## DATA SCIENCE

### Data Scientist- *Initiation* (50 hrs)

Foundations of Statistical Analysis & Machine Learning – Part 1 (25 hrs)  
Python Machine Learning Labs (25 hrs)

### Neural Networks (75 hrs)

Continuous Optimisation (25 hrs)  
Artificial Neural Networks (25 hrs)  
Deep Learning (25 hrs)

### Statistical Machine Learning – *Base* (75 hrs)

Foundations of Statistical Analysis & Machine Learning – Part 2 (40 hrs)  
Advanced Statistical Analysis & Machine Learning (35 hrs)

## DATA ENGINEER

### Data Engineer - *Initiation* (50 hrs)

Python Machine Learning Labs (25 hrs)  
Data Wrangling with SQL (25 hrs)

### Advanced Databases (100 hrs)

Hadoop & Spark (50 hrs)  
DevOps & Continuous Integration (50 hrs)  
Foundations of Statistical Analysis & Machine Learning – Part 1 (25 hrs)  
R for Big Data (25 hrs)  
Python Machine Learning Labs (25 hrs)  
Deep Learning (25 hrs)

### Big Data (75 hrs)

Relational databases for ETL & Data Warehousing (25 hrs)  
NoSQL (25 hrs)  
Data Pipeline 1 – XML & JSON (25 hrs)  
Semantic Web Technologies (25 hrs)

### Statistical Machine Learning – *Base* (75 hrs)

Amazon AWS (50 hrs)  
Azure (25 hrs)  
Cybersecurity (25 hrs)

### Big Data (75 hrs)

Hadoop & Spark (50 hrs)  
Data Pipeline – Part 2 (25 hrs)

### Statistical Machine Learning – *Advanced* (150 hrs)

Foundations of Statistical Analysis & Machine Learning – Part 2 (40 hrs)  
Advanced Statistical Analysis & Machine Learning (35 hrs)  
Statistical Analysis of Massive and High Dimensional Data (25 hrs)  
Time-Series Analysis (25 hrs)  
Survival Analysis (25 hrs)

# FRAIS DE FORMATION

Éligible aux mécanismes de financement de la formation professionnelle en France (CPF, Transitions Pro, Dispositifs d'Aides Régionales, etc.)

Les frais de formation se font en fonction de la volumétrie choisie.

**Tarif horaire HT : 25€/heure**

Exemples:

1 250€ HT  50 heures

1 875€ HT  75 heures

2 500€ HT  100 heures

3 750€ HT  150 heures

5 000€ HT  200 heures



# CAMPUS NICE SOPHIA-ANTIPOLIS

Siège

Data ScienceTech Institute

Les Templiers, 950 Route des Colles  
Sophia-Antipolis  
06410 Biot, France

Tel : 04 89 41 29 44

# CAMPUS PARIS

Data ScienceTech Institute

4 Rue de la Collégiale  
75005 Paris, France

Tel : 04 89 41 29 44



/datasciencetech



@datasciencemsc



contact@dsti.institute



@datasciencetech



[www.datasciencetech.institute](http://www.datasciencetech.institute)

Données mises à jour en mai 2022