

## Agenda

- 1) Le parcours *Univers et Particules* tel qu'il a été adopté.
  - 2) Le module *Data Science*. Déjà construit → syllabus.
  - 3) Le module *Instruments & Methods*.
  - 4) Le module *Science*.
  - 5) Le module M1 d'initiation *Univers et Particules*.
- O) Objectif de la réunion : 1/ passer en revue les enseignements, 2/ distinguer un programme pour chaque UE, 3/ commencer à identifier les enseignants possibles.

# 1) Parcours Univers et Particules

## Master 2 Univers et Particules (UP) : 300h - 60 ECTS

Data Science : 90h (12 ECTS)		Instruments & Methods : 50h (6 ECTS)	
Algorithms and Language (OO)	25h	Detector Physics	24h
Data Mining	25h	Instrumental Projects	26h
Statistics	20h		
Machine Learning	20h		
Science : 160h (24 ECTS)			
Symmetries in Physics, Group Theory ;		Introduction to Particle Physics	40h
Quantum Field Theory, Quantum chromodynamics, Quark gluon Plasma			40h
From Electro-Magnetism to the Electro-Weak Standard Model			40h
General relativity, astroparticles and cosmology			40h
Internship (+english skills - 18 ECTS)			

*Des données à la connaissance*

Triptyque : savoirs fondamentaux (Science), méthodologie (Instruments & Methods) & expertises immédiatement transférables en métier (Data Science).

Double objectif: préparer à la thèse d'Université en physique des hautes énergies et cosmologie / préparer aux métiers de *data scientists*.

## Diplôme Universitaire Data Scientist

### Présentation

- Formation dispensée par l'Université Clermont Auvergne.
- Durée de la formation : 1 an, volume horaire : 124h.
- Accessible en formation initiale (BAC+4) et en formation continue.
- Lieu de formation : UCA, campus des Cézeaux, Aubière. Le module « Architecture, Data and Computing » est dispensé au Centre de Calcul de l'IN2P3, à Villeurbanne.
- Responsable pédagogique : Julien Donini (Professeur à l'UCA, Laboratoire : LPC).

### Contexte

Dans de nombreux domaines scientifiques (physique, neurosciences, génétique, climatologie,...) ainsi que dans le monde de l'entreprise (télécoms, intelligence artificielle, économie, finances, ...) il devient de plus en plus nécessaire de pouvoir manipuler de grandes masses de données à l'aide d'infrastructures informatiques solides et rapides et de savoir les exploiter avec des algorithmes statistiques avancés. L'ensemble des méthodes employées, connues généralement sous les dénominations de « fouille de données » (data-mining) ou « datamasse » (big data), vise à extraire de ces systèmes complexes une information compacte permettant, entre autres, la visualisation, la classification et la modélisation des données. Des métiers spécifiques sont apparus au cours des dernières années autour de ces thématiques, et la demande d'experts en traitement de données (data scientist) est en pleine croissance.

### Objectifs

La formation proposée par ce Diplôme Universitaire permet à des professionnels et à des étudiants sortant d'une formation initiale de niveau Master d'acquérir les compétences nécessaires en informatique, statistiques, machine learning et fouilles de données, afin d'appréhender les problématiques liées au « big data » aujourd'hui.

Ce Diplôme Universitaire offre des débouchés dans les secteurs suivants :

- Métiers de Data Scientist (data analyst, data miner)
- Chargé d'études statistiques
- Chargé d'études prospectives et d'optimisation
- Analyste en intelligence socio-économique
- Responsable gestion et analyse de données
- Développeur analyste-programmeur
- Métiers de la Recherche

### Organisation de la formation

Les enseignements sont dispensés par des enseignants-chercheurs et des chercheurs du Laboratoire de Physique de Clermont (LPC) et du Laboratoire d'Informatique, de Modélisation et d'Optimisation des Systèmes (LIMOS). La formation bénéficie de l'implication de plusieurs de ses intervenants dans des expériences scientifiques internationales devant gérer et analyser de grandes masses de données telles que le grand collisionneur de hadrons LHC au CERN ou le télescope LSST au Chili. Un partenaire naturel sur lequel s'adosse la formation est le Centre de Calcul de l'Institut National de Physique Nucléaire et de Physique des Particules (CC-IN2P3) qui propose une expertise et un savoir-faire technologique, en particulier dans le domaine du calcul intensif sur grille informatique (grid computing).

La formation repose sur les 5 modules complémentaires suivants :

- 1) Advanced Computing : 25h
- 2) Advanced Statistics : 20h
- 3) Machine Learning : 20h
- 4) Data Mining and Big Data : 25h
- 5) Architecture, Data and Computing : 34h

Les modules 1-4 sont mutualisés avec Master Physique Fondamentale et Applications. Dans le calcul du coût du DU les heures communes au Master sont comptabilisées en HCC.

Les modules 3-5 peuvent être proposés séparément de façon modulaire à la formation continue.

### Public concerné

- Mode d'accès : formation initiale et formation continue.

- Bénéficiaires : Cette formation s'adresse aux étudiants ayant un niveau BAC+4 dans une formation scientifique ainsi qu'à des salariés d'entreprises (par ex : cadres Michelin) souhaitant acquérir des techniques de calcul statistique et de traitement de données.

- Compétences globales visées :

- Savoir concevoir des programmes concurrents.
- Maîtrise de concepts statistiques avancés.
- Mise en pratique de méthodes d'analyse prédictive et d'optimisation.
- Implémentation d'algorithmes d'apprentissage supervisé et non supervisé.
- Utilisation de techniques modernes de calcul et de traitement des données

- Pré-requis :

Les pré-requis pour chaque modules sont détaillés ci-dessous.

## 2) Module Data Science

---

Le dossier du diplôme universitaire de Data Scientist est introduit. Les tarifications sont établies en collaboration avec l'UCA. Il sera discuté le 09 mai en CFVU et examiné en juin par ce même conseil.

Nous avons toutefois reçu l'autorisation d'en faire la publicité du point de vue de la formation continue.

Les partenaires sont le LPC, le Master PFA, le LIMOS et le CC-in2p3.

Le porteur de projet est Julien Donini.

Les programmes pédagogiques des cinq UE sont établis et sont reportés dans les diapositives suivantes.

## 2) Module Data Science

Module 1		Advanced Computing
Responsable de Module :		
Présentation	Description et objectifs	Ce cours aborde la programmation concurrente au travers des exemples pratiques utilisant le langage Go.
	Compétences visées	- Comprendre les méthodes et architectures de programmation concurrente. - Savoir concevoir des programmes concurrents. - Programmation à l'aide du langage Go.
	Pré-requis	Notions d'architecture et de programmation. Connaissance de bases d'un langage informatique.
	Formation	Lieu : UCA, Campus Cézeaux Durée : 25h Date : Contact : Effectifs : Formation continue : oui
Contenu	Programme détaillé	<p>1) Introduction to concurrent programming</p> <ul style="list-style-type: none"> <li>- definition of concurrent programming and parallel programming</li> <li>- introduction to Go's concurrent execution model (CSP)</li> <li>- introduction to common issues with concurrent programming</li> <li>- case study of various concurrent programming architectures</li> <li>- concurrent/parallel control data structures (locks, mutexes, spinlocks, etc)</li> </ul> <p>2) Introduction to object oriented programming</p> <ul style="list-style-type: none"> <li>- concurrent algorithms (sort, tree-walking)</li> </ul> <p>Introduction to object oriented programming</p> <ul style="list-style-type: none"> <li>- 3 pillars of OOP (encapsulation, inheritance and polymorphism)</li> <li>- OOP in Go (interfaces and "duck-typing")</li> </ul> <p>3) Development good practices</p> <ul style="list-style-type: none"> <li>- distributed version control systems (git)</li> <li>- build methodologies, reproducible builds and results (make)</li> <li>- documenting code (godoc)</li> <li>- tests, unit tests, regressions tests and integration tests (gotest)</li> <li>- continuous integration, nightlies</li> </ul>
	Volume horaire	TP : 25h
	Équipe enseignante	1 ITA CNRS du LPC.
	Ressources et moyens	

Module 2		Advanced Statistics
Responsable de Module :		
Présentation	Description et objectifs	Cours et exercices pratiques permettant d'appréhender les concepts statistiques essentiels au 'data scientist'.
	Compétences visées	- Maîtriser des concepts statistiques avancés. - Analyser des données à l'aide de méthodes numériques. - Appliquer des méthodes d'analyse prédictive. - Construire des modèles statistiques. - Savoir programmer avec le logiciel R.
	Pré-requis	Notions de probabilités. Connaissances mathématiques (niveau Licence).
	Formation	Lieu : UCA, Campus Cézeaux Durée : 20h Date : Contact : Effectifs : Formation continue : oui
Contenu	Programme détaillé	<ul style="list-style-type: none"> <li>- Basics on theory of probabilities</li> <li>- Properties of distributions</li> <li>- Random variables distributions</li> <li>- Samples and statistical properties of samples</li> <li>- Likelihood</li> <li>- Parameter estimation</li> <li>- Hypothesis testing</li> <li>- Confidence level and confidence intervals</li> <li>- Bayesian inference</li> <li>- R statistical software</li> </ul>
	Volume horaire	CM : 15h TD : 5h
	Équipe enseignante	1 enseignant-chercheur UCA du LPC.
	Ressources et moyens	

## 2) Module Data Science

Module 3		Machine Learning
Responsable de Module :		
Présentation	Description et objectifs	Introduction aux techniques multi-variables et aux méthodes d'apprentissage supervisé (Machine Learning).
	Compétences visées	- Appliquer des méthodes de classification et de clustering. - Implémentation d'algorithmes d'apprentissage supervisé. - Concevoir des méthodes d'analyse prédictive et d'optimisation. - Utiliser un algorithme de Deep Learning.
	Pré-requis	Notions de probabilités et de statistiques. Connaissances mathématiques (niveau Licence).
	Formation	Lieu : UCA, Campus Cézeaux Durée : 20h Date : Contact : Effectifs : Formation continue : oui
Contenu	Programme détaillé	- Introduction to machine learning - Discriminating variables, regression, classification, clustering - Dimensionality reduction, LDA, PCA - K nearest neighbour classifier - Supervised machine learning - Decision trees, neural networks - Deep learning
	Volume horaire	CM : 10h TP : 10h
	Équipe enseignante	1 enseignant-chercheur UCA (LPC) 1 intervenant extérieur (Yandex)
	Ressources et moyens	

Module 4		Data mining and big data
Responsable de Module :		
Présentation	Description et objectifs	Introduction aux méthodes de fouille de données et aux techniques d'apprentissage non supervisé.
	Compétences visées	- Savoir évaluer les méthodes de fouille de données. - Savoir utiliser les techniques de fouille de données. - Maîtriser une technique de classification non supervisée. - Maîtriser une technique de recherche de règles d'association. - Connaître les différentes approches de fouille de données complexes.
	Pré-requis	
	Formation	Lieu : UCA, Campus Cézeaux Durée : 25h Date : Contact : Effectifs : Formation continue : oui
Contenu	Programme détaillé	Data Mining - Introduction à l'extraction des connaissances. - Prétraitement des données. - Environnement méthodologique de la fouille de données. - Règles d'association. - Classification non supervisée. - Fouille de données complexes. - Panorama des outils de fouille de données (WEKA, SAS, ...). Big data - Apprentissage et Passage à l'échelle - Clustering et Passage à l'échelle - Fouille de masse de données - Visualisation de masse de données - Applications - Outils : Apache MAHOUT, SHARK, Apache SPARK
	Volume horaire	TD : 25h
	Équipe enseignante	1 enseignant-chercheur UCA du LIMOS.
	Ressources et moyens	

## 2) Module Data Science

Module 5		Architecture, data and computing
Responsable de Module :		
Présentation	Description et objectifs	Formation aux techniques modernes de calcul scientifique et de traitement des données dispensée par le Centre de Calcul de l'IN2P3 (cc.in2p3.fr).
	Compétences visées	<ul style="list-style-type: none"> <li>- Comprendre les problématiques liées au calcul (architectures modernes, calcul sur la grille et dans le cloud)</li> <li>- Savoir utiliser une grille de calcul.</li> <li>- Savoir choisir le stockage le plus adapté : système de fichiers, stockage dans le cloud et bases de données (SQL et NoSQL).</li> <li>- Programmation avec le langage Python,.</li> <li>- Utilisation d'un notebook de type Jupyter pour mener une analyse de données.</li> </ul>
	Pré-requis	Connaissance minimale d'un système d'exploitation (Linux de préférence) et de la ligne de commande. Connaissance minimale d'un langage de programmation (écriture du code, compilation et exécution)
	Formation	Lieu : Centre de Calcul de l'IN2P3 (Villeurbanne) Durée : 34h, formation sur 5 jours. Date : Contact : Effectifs : Formation continue : oui
Contenu	Programme détaillé	1) Mise à niveau - Linux, ligne de commande et réseau IP. 2) Python et "notebooks" - introduction au langage Python, et aux "notebooks » - Utilisation d'un "notebook" Jupyter 3) Calcul - architectures de calcul - virtualisation et calcul distribué : cloud et grille de calcul 4) Stockage - système de fichiers - bases de données relationnelles et non relationnelles 5) Visite de la salle des machines du Centre de Calcul.
	Volume horaire	CM : 16h TP : 18
	Équipe enseignante	8 intervenants ITA du CCIN2P3.
	Ressources et moyens	Ressources informatiques du CCIN2P3. Déplacement et séjour (1 semaine) à Villeurbanne.

Les modules 1, 3 et 5 seront placés dès cette année en Formation Continue.

Contraintes d'EdT : les modules 1 et 3 seront donnés sur une semaine.

Messages:

- La FC devrait nous permettre de construire une liste de partenaires pour l'accueil en stage de nos étudiants.
- La FC fournit des ressources propres qui doivent servir le Master.

### 3) Instruments & Methods

#### Detector Physics:

- .) dE/dx Bethe-Bloch formula.
- .) Ionisation detectors.
- .) Scintillation.
- .) Silicon detectors.
- .) Tracking
- .) Calorimetry
- .) Nuclear Physics detectors (n,  $\gamma$ ).
- .) CCD-like cameras

Applications by examples: Aleph, Atlas.

#### **Projects:** low-level communications.

Les projets s'organisent autour de huit séances hebdomadaires de deux heures et une semaine complète.

David Calvet est responsable de cet enseignement.

Il a introduit une demande pour l'achat de PC auprès de l'Eupi (seule demande du Master PFA à ma connaissance) et utilisera des modules Arduino.

L'enseignement commencera après le cours d'Info.

### Symmetries (20h):

#### Introd. to Particle Physics(20h):

- .) Useful concepts: antimatter, decay width, cross-sections, quantum interactions
- .) Quarks, hadrons, jets
- .) Parity violation
- .) Helicity and chirality
- .) V-A cc interaction.
- .) Neutral currents.
- .) Cabibbo model, Intro to CKM.
- .) Weak mixings, CP violation
- .) Introduction to Standard Model.

- .) Voir document attaché à l'agenda.
- .) Introduction aux symétries en mécanique classique et quantique : contraintes sur les lois physiques et grandeurs conservées
- .) Théorème de Noether
- .) Outils mathématiques : théorie des groupes et de la représentation (groupes finis, groupes de Lie, représentations réductibles et irréductibles, lemmes de Schur, critères de recherche des représentations irréductibles, classification de Cartan, poids, décomposition à la Clebsch-Gordan)
- .) Applications physiques en mécanique quantique
- .) Théorème de Wigner
- .) Structures cristallines
- .) Molécules (orbitales, vibration...)
- .) Règles de sélection atomique
- .) Multiplets de particules
- .) Atome d'hydrogène.

### General relativity (20h):

- .) Special relativity wrap-up: Lorentz transformation, covariant formalism...
- .) Riemann space formalism.
- .) Equivalence principle, Einstein equations, Hilbert lagrangian, Planck scale.
- .) Horizons: Schwarzschild metric, black holes, horizon and inflation.

### Astroparticles and Cosmology

Proposition de programme de la partie « Astroparticles and Cosmology » (20 heures) du module *General relativity, astroparticles and cosmology* – élaboration à partir des références suivantes :

- I. *Cosmologie*, J. Rich, Vuibert, 2010.
- II. *Theoretical concepts in physics*, M. Longair, Cambridge University Press, 2003.
- III. *An introduction to modern cosmology*, A. Liddle, John Wiley & Sons, 2015.
- IV. *Lectures on Astronomy, Astrophysics and Cosmology*, L. A. Anchordoqui, arXiv:0706.1988 [physics.ed-ph], 15 Jun 2016.

Cette proposition, certainement trop ambitieuse pour 20 heures de cours, affiche clairement un contenu avant tout « cosmologie ». Le domaine des « astroparticules » est introduit comme un outil en relation avec la cosmologie, avec le biais de ne pas connecter les astroparticules avec les questions astrophysiques sous-jacentes (sources de production des rayons cosmiques, mécanismes de production et d'accélération, etc.).

- 1. Introduction to observational cosmology**
  - Astronomical objects and basic astrophysics
  - Distances, redshift and Hubble law
- 2. Standard model of cosmology ( $\Lambda$ CDM)**
  - Friedmann-Lemaître equations
  - Cosmological parameters of an homogeneous Universe (curvature, radiation, matter and cosmological constant)
  - Observational constraints (flatness, dark matter and dark energy)
  - Age and size of the observable Universe (concept of "horizon")
- 3. Thermal history of the Universe**
  - Particle physics phase transitions
  - Big bang nucleosynthesis (BBN)
  - Recombination and cosmic microwave background (CMB)
- 4. Inhomogeneous Universe**
  - Structure formation
  - Observable anisotropies and cosmological constraints (baryon acoustic oscillations and CMB fluctuations)
  - Open questions in cosmology and inflation solution (horizon problem, flatness problem and initial fluctuations)
- 5. Astroparticles and multi-messenger astronomy**
  - From radio to gamma astronomy
  - Cosmic rays, cosmic neutrinos and dark matter searches
  - Gravitational waves

De P. Rosnet —◆—

Prérequis : relativité générale, astronomie et astrophysique, thermodynamique et physique statistique, physique nucléaire et des particules

### Quantum Field Theory (20h):

Voir document attaché

### Strong interactions (20h):

- .) Modèle des quarks (avec intro  $SU(3)_F$ )
- .) Introduction de la couleur
- .) hadrons et leurs désintégrations
- .) Les outils d'étude de la structure des hadrons : diffusion élastique de leptons et facteurs de forme, diffusion inélastique de leptons et le modèle des partons
- .) Phénoménologie de l'interaction forte
- .) Lagrangien de QCD, constante de couplage et liberté asymptotique
  
- .) Potentiel quark-quark et confinement
- .) Diagramme de phase de QCD et plasma de quarks-gluons
- .) Collisions proton-proton, modèle de Glauber et collisions d'ions lourds

## 4) Science

### From e.m. to the electroweak SM (40h):

10h

- .) Abelian gauge theory and QED.
- .) QED examples and calculations [e.g.  $(g-2)$ ].
- .) Non-abelian gauge theory and QCD.
- .) Propagator corrections, renormalisation.

15h

- .) Electroweak unification : charged and neutral currents.
- .) Full calculation of the Z width, number of active neutrinos.
- .) Spontaneous symmetry breaking : BEH mechanism.
- .) Electroweak precision tests.

.) Flavour Physics (5h).

.) Neutrino Physics (5h).

.) BSM

## 5) Introduction en M1

---

Un module de 50 h

Préexistaient:

- .) Cours de 20 h dans les fondamentaux du M1.
- .) 2 UE de 50 h: MQR (20h), Th. des collisions (15h), neutronique (25h), Physique des détecteurs (15h).
- .) 2 TP: analyse de données avec le détecteur Aleph (8h), muon lifetime (10h).

Notes:

- La MQR est dans les fondamentaux de la nouvelle maquette.
- Le programme du M2 doit être self-contained en raison du recrutement d'étudiants étrangers en M2. On a donc une certaine liberté ici.
- Une visite du CERN aurait un intérêt.

## 5) Introduction en M1

---

Proposition pour les 50 h:

.) Cours de 15h d'introduction à la physique des particules.

.) Cours de 10h Théories des collisions.

.) Travaux pratiques: temps de vie du muon (10h), analyse de données (10h).

.) Visite du CERN et cours sur les accélérateurs (5h)