

**SOSC 2018 Second  
International PhD School  
Open Science Cloud**



**Report of Contributions**

Contribution ID: 0

Type: **not specified**

## **School on Open Science Cloud - Welcome and Introduction**

*Monday, September 17, 2018 10:00 AM (20 minutes)*

**Presenter:** SPIGA, Daniele (PG)

**Session Classification:** Introduction

Contribution ID: 2

Type: **not specified**

## Predictive Models for Thermal Modelling

*Monday, September 17, 2018 11:50 AM (1 hour)*

We hear it almost every day but we barely recognize, it's the noisy fan of our laptop. We feel it with our hands in the summer, it's our mobile getting too hot. It's the heat dissipated by electronics devices. Digital processing elements, the heart of our smartphone, laptop, workstation and super-computers dissipate power for flipping bits of information, this power increases the temperature of the silicon. The heat must be removed to keep the electronics safe.

In the Multitherman ERC advance project we studies techniques for extract physically valid compact models directly from the final device, and to combine these with optimization strategies preserve the working temperature of the processing element. In this presentation I will give the basic knowledge on the problem an overview of the recent advances in the thermal modelling of real electronics devices.

**Primary author:** Prof. BARTOLINI, Andrea (Università di Bologna)

**Presenter:** Prof. BARTOLINI, Andrea (Università di Bologna)

**Session Classification:** Introduction

Contribution ID: 3

Type: **not specified**

## **Predictive models with Machine and Deep Learning: a scientific view**

*Monday, September 17, 2018 10:20 AM (1 hour)*

An overview of (selected) use-cases for {Machine,Deep} Learning in various scientific disciplines is presented. Peculiarities of - and similarities across - disciplines are explored. Possible cross-fertilisations and synergic approaches, as well as the value of training paths, is also discussed.

**Primary author:** BONACORSI, Daniele (BO)

**Presenter:** BONACORSI, Daniele (BO)

**Session Classification:** Introduction

Contribution ID: 4

Type: **not specified**

## Python Coding session 1

*Monday, September 17, 2018 2:00 PM (1h 35m)*

Material is available here <https://github.com/calde12/SOSC2018-PYTHON-HANDSON>

**Primary author:** Dr CALDERAN, Stefano (Oval Money)

**Presenter:** Dr CALDERAN, Stefano (Oval Money)

**Session Classification:** Hands-on - Introductory Python Programming

Contribution ID: 5

Type: **not specified**

## **Advanced Python Coding (Speaker TBD)**

Contribution ID: 6

Type: **not specified**

## Python Coding session 2

Contribution ID: 7

Type: **not specified**

## Introduction to Big Data

*Tuesday, September 18, 2018 9:00 AM (1 hour)*

In this presentation we will introduce the concept and distinctive characteristics of Big Data. Examples taken from both scientific and commercial realms will be provided and will illustrate how BigData analytics is a key enabler for social business, but also science can benefit from using the same techniques. A quick overview of BigData computing infrastructures, in particular their relation to IoT, Cloud, Edge and Fog computing paradigms will be presented.

**Primary author:** CESINI, Daniele (CNAF)

**Presenter:** CESINI, Daniele (CNAF)

**Session Classification:** Big Data and Machine Learning



Contribution ID: 8

Type: **not specified**

# Introduction to Machine Learning and Deep Learning

*Tuesday, September 18, 2018 10:00 AM (1 hour)*

From self-driving cars to voice-activated intelligent assistants and robots reading and mimicking human emotions, machine learning and deep learning are not only revolutionizing several industry sectors but they are slowly and steadily becoming part of our daily lives.

This talk will provide an overview of deep learning, namely of modern, multi-layered neural networks trained on big data. In the talk the basic theory of deep learning and of neural networks' training and optimization will be introduced. Then, the main types of deep learning models will be presented, together with some relevant applications in the fields of computer vision, robotics and natural language process.

**Primary author:** Prof. RICCI, Elisa (Università di Trento)

**Presenter:** Prof. RICCI, Elisa (Università di Trento)

**Session Classification:** Big Data and Machine Learning

Contribution ID: 9

Type: **not specified**

## Data Science: state of the art

*Tuesday, September 18, 2018 11:30 AM (1 hour)*

The Data Science is an rapidly growing field of Information Technology. In this talk we'll cover its origin, the connection to other IT fields and cover details of Data Science fields from data acquisition, processing to building Machine Learning models. We'll cover the tools, technologies and techniques a Data Scientist use in their daily activities. We'll also discuss algorithms and fine-tuning tricks which will allow you to train and deliver world class Machine Learning models.

**Primary author:** Dr KUZNETSOV, Valentin (Cornell University (US))

**Presenter:** Dr KUZNETSOV, Valentin (Cornell University (US))

**Session Classification:** Big Data and Machine Learning

Contribution ID: 11

Type: **not specified**

## Hands-on: Deploy ML Frameworks 101

*Tuesday, September 18, 2018 2:00 PM (1h 20m)*

In this Hands-On session we will start from bare node with deploy all necessary tools to perform Machine Learning and Data Science studies. The session will require basic knowledge of Linux operating system and UNIX shell. We'll download, deploy and install Anaconda environment where we'll install various set of Python and R based packages. We'll work with a simple Iris dataset to explore our data.

**Primary author:** Dr KUZNESTOV, Valentin

**Presenter:** Dr KUZNESTOV, Valentin

**Session Classification:** Hands-on

Contribution ID: 12

Type: **not specified**

## **Computational and Statistical Learning - (Speaker TBD)**

Contribution ID: **13**

Type: **not specified**

## **Hands-on 2**

Contribution ID: 14

Type: **not specified**

## **Models and Algorithms in HEP**

*Wednesday, September 19, 2018 11:30 AM (1 hour)*

**Primary authors:** ZANETTI, Marco (PD); PIERINI, Maurizio (CERN)

**Presenters:** ZANETTI, Marco (PD); PIERINI, Maurizio (CERN)

**Session Classification:** Algorithms and Models

Contribution ID: 15

Type: **not specified**

## Models and Algorithms in Medical Imaging

*Wednesday, September 19, 2018 9:00 AM (1 hour)*

Medical Imaging techniques rely on different physical principles to obtain signals from the human body. Medical images are more than pictures, they are data, and they can be explored through image analysis techniques to go well beyond the mere visual inspection by radiologists. Image processing and data mining techniques are used to extract useful information from medical images to assist clinicians in the diagnosis and follow up of diseases, and for the planning, guidance and monitoring of treatments. Significant progress has been made in the last few decades by applying data mining techniques to the problems of image segmentation, image registration, feature extraction and interpretation, relying on the typical instruments of the artificial intelligence, e.g. data driven models built with machine learning techniques. The latter are rapidly leaving space to deep learning applications, thanks to the recent availability of suitable computing power.

An overview on the data mining approaches used in Medical Imaging will be provided, with realistic examples in the field of neuroimaging with magnetic resonance. The problems of data harmonisation and confounding parameters in multicenter studies will be highlighted, with particular reference to the ARIANNA project, which has built an interdisciplinary research platform for neuroimaging-based studies of Autism Spectrum Disorders.

**Primary author:** RETICO, Alessandra (PI)

**Presenter:** RETICO, Alessandra (PI)

**Session Classification:** Algorithms and Models

Contribution ID: 16

Type: **not specified**

## **The Ophidia project: towards a High Performance Data Analytics and Machine Learning framework for climate change**

*Wednesday, September 19, 2018 10:00 AM (1 hour)*

**Primary author:** Dr FIORE, Sandro (Centro Euro-Mediterraneo sui Cambiamenti Climatici)

**Presenter:** Dr FIORE, Sandro (Centro Euro-Mediterraneo sui Cambiamenti Climatici)

**Session Classification:** Algorithms and Models



Contribution ID: 17

Type: **not specified**

## **Hands-on: Data analytics with Ophidia/ECAS**

*Wednesday, September 19, 2018 2:00 PM (1h 20m)*

**Primary author:** Dr FIORE, Sandro (CMCC)

**Presenter:** Dr FIORE, Sandro (CMCC)

**Session Classification:** Hands-on

Contribution ID: **18**

Type: **not specified**

## **Coffee Seminar: Machine Learning for Particle Physics**

*Wednesday, September 19, 2018 3:35 PM (55 minutes)*

**Primary author:** Dr PIERINI, Maurizio (CERN)

**Presenter:** Dr PIERINI, Maurizio (CERN)

**Session Classification:** Coffee Seminar

Contribution ID: **19**

Type: **not specified**

## **Hands-on 2**

Contribution ID: 20

Type: **not specified**

## **Deep Learning and Machine Learning in hybrid clouds**

*Thursday, September 20, 2018 9:00 AM (1 hour)*

**Primary author:** SALOMONI, Davide (CNAF)

**Presenter:** SALOMONI, Davide (CNAF)

**Session Classification:** Infrastructures

Contribution ID: 21

Type: **not specified**

## Big Data at CERN

*Thursday, September 20, 2018 10:00 AM (1 hour)*

The LHC experiments continue to produce a wealth of valuable High Energy Physics data, which offer numerous possibilities for new discoveries. The IT Department at CERN provides Hadoop and Spark services and works closely with the scientific communities in their quest to analyze and understand these vast amounts of physics and infrastructure data. The number of CERN teams using these services for their systems has grown significantly over the past years, since Big Data technologies -such as Apache Spark- show great potential in speeding up their existing workloads. The most significant systems include the CMS Data Reduction Facility which aims to reduce 1 PB of data produced by the CMS Experiment to 1 TB of reusable data for physics analysis through Spark, the Next CERN Accelerator Logging Service (NXCAL) which will perform online and offline analysis over the data acquired from each of the 20,000 devices that monitor the CERN accelerator complex, as well as the monitoring system for the CERN Data Center and the Worldwide LHC Computing Grid (WLCG) which consists of more than 170 different computing centers in 42 countries. This talk will provide an overview of the current infrastructure based on Spark and other key components of the Hadoop ecosystem, the active use cases on big data analytics from various CERN communities, as well as the challenges in the available data sources and their architecture.

**Primary author:** Dr MOTESNITSALIS, Vaggelis (CERN)

**Presenter:** Dr MOTESNITSALIS, Vaggelis (CERN)

**Session Classification:** Infrastructures

Contribution ID: 22

Type: **not specified**

## **T-Systems, Open Telekom Cloud**

*Thursday, September 20, 2018 11:30 AM (1 hour)*

**Primary author:** Dr DE LA MAR, Jurry (T-System)

**Presenter:** Dr DE LA MAR, Jurry (T-System)

**Session Classification:** Infrastructures

Contribution ID: 23

Type: **not specified**

## **Hands on: Automated deployment of BigData Cluster in the Cloud**

*Thursday, September 20, 2018 2:00 PM (1h 20m)*

This session will include the usage of Dynamic On-Demand Analysis Service (DODAS) and as such will include:

Containers Orchestration (Mesos)

Software Applications & Dependency description (TOSCA);

Cloud PaaS Orchestration (INDIGO-PaaS)

**Primary authors:** SPIGA, Daniele (PG); CIANGOTTINI, Diego (PG); TRACOLLI, Mirco (PG)

**Presenters:** SPIGA, Daniele (PG); CIANGOTTINI, Diego (PG); TRACOLLI, Mirco (PG)

**Session Classification:** Hands-on

Contribution ID: 24

Type: **not specified**

## Coffee Seminar: Becoming a Data Scientist

*Thursday, September 20, 2018 3:35 PM (55 minutes)*

In this talk we'll cover materials and knowledge base of how to become a world class Data Scientists. We'll use real kaggle competition dataset and build for it a set of ML models. During this session we'll introduce advanced concepts of ML training, like feature embedding, data transformation and ML model fine-tuning. We'll discuss how to work large datasets, techniques to avoid RAM limitations problems, and how to create an ensemble of multiple ML models. We'll use XGBoost library and Keras ML framework to build our models and embeddings matrix. Even though the materials of this session is quite advance we expect students to listen and follow up with introduced concepts.

**Primary author:** Dr KUZNETSOV, Valentin (Cornell University (US))

**Presenter:** Dr KUZNETSOV, Valentin (Cornell University (US))

**Session Classification:** Coffee Seminar



Contribution ID: 25

Type: **not specified**

## Hands-on 2

Contribution ID: 26

Type: **not specified**

## Research & Business - NVIDIA

*Friday, September 21, 2018 11:30 AM (1 hour)*

The convergence of supercomputing and AI in a Post-Moore's Law World

The world is facing many very large challenges in many fields, including energy, climate, and biology. High-performance computing can help provide the answers through large-scale simulation, but ground-breaking simulations are incredibly computationally expensive.

Unfortunately, Moore law is slowly stopping and there is no single technology that might revamp it in the near future. At the same time new approaches to scientific computing are rising and they are changing the paradigm to model natural phenomena or analyze data generated by large scientific instruments. Now it is the era of the of AI+HPC.

Here will be briefly presented the current applications that successfully take advantage of the hybrid HPC+AI approach and what is expected to happen in the future.

**Primary author:** Dr ALTOE', Piero (NVIDIA)

**Presenter:** Dr ALTOE', Piero (NVIDIA)

**Session Classification:** From research to business

Contribution ID: 27

Type: **not specified**

## Research & Business - OPENLab - IT Technology Transfer at CERN

*Friday, September 21, 2018 9:00 AM (1 hour)*

CERN openlab as a model for R&D and Technology Transfer at CERN

**Abstract:** CERN openlab was created in 2001 as a way of establishing a new collaboration channel between CERN and ICT industry.

Over the past 17 years, it has become a reference for joint R&D and technology transfer at CERN in support of the future computing and data requirements of the increasingly challenging LHC research programme. This talk will show how CERN openlab works and what are the elements that have allowed successful public-private partnership to be established in as diverse domains as data acquisition, storage, computing platforms and software, cloud computing, networks or machine learning. The current technological challenges and the directions of investigations will be briefly explained as well as the importance of communication and education to support ambitious R&D programmes.

**Primary author:** Dr DI MEGLIO, Alberto (CERN)

**Presenter:** Dr DI MEGLIO, Alberto (CERN)

**Session Classification:** From research to business

Contribution ID: 28

Type: **not specified**

## Applications to Finance and Insurance

*Friday, September 21, 2018 10:00 AM (1 hour)*

Some applications of big data analytics techniques to the insurance business

Some examples are provided to illustrate how the insurance industry is currently facing the new paradigm of big data analytics. Methodologies for analyzing telematic car driving data are illustrated. In particular, it is shown how pattern recognition and machine learning techniques can be used to derive predictive models useful for car insurance pricing. Moreover, a machine learning application to individual claims reserving is presented. Classification and regression trees techniques are used to realize claim watching, i.e. to predict and control the cost development process of individual claims.

**Primary author:** Prof. MORICONI, Franco (University of Perugia and ALEF Advanced Laboratory Economics and Finance srl)

**Presenter:** Prof. MORICONI, Franco (University of Perugia and ALEF Advanced Laboratory Economics and Finance srl)

**Session Classification:** From research to business

Contribution ID: 29

Type: **not specified**

## **Follow-up projects & collaboration opportunities**

*Friday, September 21, 2018 12:30 PM (1 hour)*

**Presenter:** SALOMONI, Davide (CNAF)

**Session Classification:** From research to business

Contribution ID: 30

Type: **not specified**

## Unsupervised Learning

*Tuesday, September 18, 2018 3:35 PM (55 minutes)*

I introduce the basic concepts of unsupervised learning, highlighting the differences with supervised learning. The focus will be on three of the main tasks addressed by unsupervised learning: cluster analysis, anomaly detection and dimensional reduction. For each of them, I describe the problem in general terms, the ways to approach it, and some of the most popular algorithms to solve it.

**Primary author:** Prof. DE SIMONE, Andrea (sissa)

**Presenter:** Prof. DE SIMONE, Andrea (sissa)

**Session Classification:** Coffee Seminar

Contribution ID: 31

Type: **not specified**

## **Hands on: Automated deployment of BigData Cluster in the Cloud**

*Thursday, September 20, 2018 4:30 PM (1h 30m)*

**Primary authors:** SPIGA, Daniele (PG); CIANGOTTINI, Diego (PG); TRACOLLI, Mirco (PG)

**Presenters:** SPIGA, Daniele (PG); CIANGOTTINI, Diego (PG); TRACOLLI, Mirco (PG)

**Session Classification:** Hands-on

Contribution ID: 32

Type: **not specified**

## Hands-on: Deploy ML Frameworks 101

*Tuesday, September 18, 2018 4:30 PM (1h 30m)*

This Hands-On session is dedicated to train your skills in area of Machine Learning (aka ML 101 course). We'll start with an Iris dataset and build our first ML model for it. Then we'll gradually expand our model to include various ML techniques, like ensemble learning, introduce concept of cross-validation and build simple Neural Network model. This session only requires basic Python programming skills and understanding of basic ML concepts.

**Primary author:** Dr KUZNETSOV, Valentin

**Session Classification:** Hands-on



Contribution ID: **33**

Type: **not specified**

## **Hands-on: The HEP Use Case**

*Wednesday, September 19, 2018 4:30 PM (1h 30m)*

Tutorial material URL: [https://www.dropbox.com/s/or43zo8imt52l0x/SOSC18\\_Tutorial.tar.gz?dl=0](https://www.dropbox.com/s/or43zo8imt52l0x/SOSC18_Tutorial.tar.gz?dl=0)

**Primary author:** Dr PIERINI, Maurizio (CERN)

**Presenter:** Dr PIERINI, Maurizio (CERN)

**Session Classification:** Hands-on

Contribution ID: 34

Type: **not specified**

## Coffe Seminar

**Presenter:** CERLINI, Paolina (CIRIAF-CRC)

Contribution ID: 35

Type: **not specified**

## ECMWF new data centre in Bologna : new opportunities for using Open Data

*Monday, September 17, 2018 3:35 PM (55 minutes)*

In the next future the new data centre of the European Centre for Medium-range Weather Forecasts (ECMWF) based in Reading U.K. will be moved to Bologna, Italy.

The building is to be delivered to ECMWF by 2019 and will host the Centre's new supercomputers, whilst the Centre's headquarters are to remain in the UK.

This centre will host one of the few global dataset of environmental data used for global numerical weather prediction as well as for climate and many other purposes that include the Climate data store (CDS) developed by ECMWF and its subcontractors within Copernicus. The use of the CDS toolbox on the top of re-analyses offers potential for implementation of data mining and deep learning algorithms. A list briefly described of topics that could be studied using new learning algorithms on ECMWF data will be given, and on the other side the 4D-Var data assimilation system used inside the ensemble method as an example of the ensemble based dynamical forecasts as used in the modern meteorological community, are presented.

**Presenter:** CERLINI, Paolina (CIRIAF-CRC)

**Session Classification:** Coffee Seminar

Contribution ID: 36

Type: **not specified**

## **Hands-on: k-means clustering, anomaly detection e PCA**

*Tuesday, September 18, 2018 6:00 PM (1 hour)*

Hands-on material : [https://github.com/de-simone/SOSC18\\_handson](https://github.com/de-simone/SOSC18_handson)

**Primary author:** Prof. DE SIMONE, Andrea (sissa)

**Presenter:** Prof. DE SIMONE, Andrea (sissa)

**Session Classification:** Hands-on