

Tools and techniques for massive data analysis

April 12-13-14, 2016 - Casalecchio di Reno, Bologna

Material: <https://goo.gl/Pq03qz>

	Tuesday, April 12	
09.00 - 09.30	Registration	
09.30 - 09.40	Introduction/logistics/agenda	<i>Giuseppe Fiameni*</i>
09.40 - 10.30	Big Data and HPC	<i>Giuseppe Fiameni</i>
10.30 - 11.00	Coffee break	
11.00 - 11.40	Brief introduction to Map-Reduce and Hadoop (Hadoop, HDFS, Yarn)	<i>Giuseppe Fiameni</i>
11.40 - 12.30	Set-up of the environment	<i>Paolo D'Onorio De Meo*</i> <i>Giuseppe Fiameni</i>
12.30 - 14.00	Lunch break	
14.00 - 15.00	Brief introduction to iPython MrJob/Hadoop Exercises	<i>Paolo D'Onorio De Meo</i> <i>Giuseppe Fiameni</i>
15.00 - 15.30	Coffee Break	
15.30 - 17.30	MrJob/Hadoop Exercises	<i>Paolo D'Onorio De Meo</i> <i>Giuseppe Fiameni</i>
17.30	End of the first day	

	Wednesday, April 13	
09.30 - 11.00	MrJob/Hadoop Exercises	<i>Paolo D'Onorio De Meo</i> <i>Giuseppe Fiameni</i>
11.00 - 11.15	Coffee break	
11.15 - 12.30	MrJob/Hadoop Exercises (NGS Use Case)	<i>Paolo D'Onorio De Meo</i> <i>Giuseppe Fiameni</i>

12.30 - 14.00	Lunch break	
14.00 - 15.00	Apache Spark	<i>Giovanni Simonini** Giorgio Pedrazzi*</i>
15.00 - 16.00	Apache Spark Examples	<i>Giovanni Simonini Giorgio Pedrazzi</i>
16.00 - 16.30	Coffee break	
16.30 - 17.30	Apache SparkSQL	<i>Giovanni Simonini Giorgio Pedrazzi</i>
17.30	End of the second day	

	Thursday, April 14	
09.00 - 09.30	Apache SparkR	<i>Giorgio Pedrazzi Giovanni Simonini Song Zhu**</i>
09.30 - 10.00	Apache Spark Exercises	<i>Giovanni Simonini Giorgio Pedrazzi Song Zhu</i>
10.00 - 12.30	Apache Spark Exercises	<i>Giovanni Simonini Giorgio Pedrazzi Song Zhu</i>
12.30 - 14.00	Lunch Break	
14.00 - 15.30	Apache Spark MLLib	<i>Giorgio Pedrazzi Song Zhu</i>
15.30 - 17.30	Apache Spark MLLib	<i>Giorgio Pedrazzi Song Zhu</i>

*Giuseppe Fiameni, Paolo D'Onorio De Meo, Giorgio Pedrazzi - **CINECA**

Giovanni Simoni, Song Zhu - **University of Modena and Reggio Emilia