Francesco Palini

linkedin.com/in/fpalini
github.com/fpalini

Experience

Research Grant Rome, Italy
Dept. of Statistical Sciences, Sapienza University of Rome 10/20 - 11/20

Winner of the grant for the research project *"Big Data Software Services for Decision Support in Precision Medicine"*, developed in collaboration with *Kazaam Lab Srl* and financed by the *European Institute for Innovation and Technology*. The activity is focused on the design and development of a platform, hosted on **Google Cloud**, that allows to analyze large PPI (Protein-Protein Interaction) networks. The data are maintained on **Neo4J** and **MongoDB** databases, while the analysis is carried out by an **Apache Spark** cluster.

Research Fellow Rome, Italy
Dept. of Statistical Sciences, Sapienza University of Rome 10/19 - 09/20

The research project *"Sviluppo, implementazione e sperimentazione di algoritmi distribuiti per l'analisi comparativa di sequenze genomiche"* is focused on the development and experimentation of distributed algorithms modelled as **MapReduce** applications, by using Apache **Hadoop** and **Spark**. The algorithms are designed to solve bioinformatic problems, with particular interest on the comparison of genomic sequences.

Teacher Rome, Italy
Dept. of Statistical Sciences, Sapienza University of Rome 03/20 - 06/20 (20h)

Teacher of the *Elaborazione distribuita di Big Data in Python* course, belonging to the *"Data Intelligence e Strategie Decisionali" (DISD)* Master. The course was focused on data analysis using the **Python** programming language, both using an **SQL**-based API to interact with an **RDBMS** and using Apache **Spark** to develop distributed algorithms for Big Data applications.

Research Grant Rome, Italy
Dept. of Statistical Sciences, Sapienza University of Rome 07/19 - 08/19

Winner of the research grant with title *"Sviluppo di un modulo software per l'accelerazione e l'ottimizzazione di procedure statistiche mediante l'uso del supercalcolo"*. The activity includes the implementation of optimized statistical software written in **R** and **C++**, with the employment of the *Terastat* supercomputing cluster.

Research Fellow Rome, Italy
Dept. of Statistical Sciences, Sapienza University of Rome 05/18 - 04/19

Involved the development of statistical software packages written in **R**, **MATLAB** and **C++**. I also worked on several bioinformatics research topics by contributing with the development and the experimentation of distributed algorithms, running on **Hadoop**, for some fundamental tasks existing in this area.

Education

Advanced School in AI (AS-AI) Rome, Italy
Institute of cognitive sciences and technologies (ISTC-CNR) (6 months) 24/07/20

Website: https://as-ai.org/

Project: "Device Alternative Interaction System (dAIsy)"

Developing a **Python** application that simplifies the interaction with vision-based devices, like **drones** or robots. Moreover, the application integrates assisted controls, by using **real-time object detection**, through an implementation of the YOLOv3 **Deep Learning** architecture.

MSc in Computer Science Rome, Italy
Sapienza University of Rome 01/18

Grade: 110/110

Thesis: "Analisi e implementazione di indicatori di progresso in Apache Pig"

Development and analysis of progress indicators for **Apache Pig**. Progress indicators are used to compute the remaining execution time for a Pig application. The main focus of the thesis was experimentation and comparison of progress indicators, analysing their pros and cons.

BSc in Computer Science Rome, Italy
Sapienza University of Rome 01/16

Grade: 110/110 with honors

Thesis: "Sistema di guida e acquisizione dati per un drone ad ala fissa"

Design and development of a control system for a fixed-wing drone. The drone was controlled by instantiating an end-to-end communication with a ground control station, represented by a laptop, by using two **Arduino** microcontrollers and **XBee** modules. The protocol used to exchange messages between the drone and the laptop was the **MAVlink** protocol.

Certifications

Coursera Certificate, R Programming 20/04/18
Department of Biostatistics, Johns Hopkins University

Online no-credit course authorized by Johns Hopkins University, taught by prof. Jeff Leek, prof. Roger Peng and prof. Bian Caffo.

Coursera Certificate, Machine Learning 13/03/18
Department of Computer Science, Standford University

Online no-credit course authorized by Standford University, taught by prof. Andrew Ng.

Oracle Certified Associate, Java SE 5/SE 6 14/12/13
Titel Pafal Group, Rome, Italy

Course organized by *Titel Pafal Group*. The exam was taken at *RTS Testing and Language Service* center.

Languages

Language: Italian Mother tongue

Language: English Intermediate

Skills

Programming Languages: Java, Python, SQL, Bash and experiences with C/C++, MATLAB, R.

IDEs: IntelliJ IDEA, PyCharm, Eclipse, Visual Studio Code, Jupyter

Build Tools: Maven, Gradle

Versioning Tools: Git, GitKraken

Frameworks: Hadoop, Spark, Spring

Programming Paradigms: MapReduce, Pregel

DBMS: MySQL, PostgreSQL, Neo4J, MongoDB

Cloud Platforms: Google Cloud, GARR Cloud

Some Libraries: numpy, pandas, matplotlib, scikit-learn, keras, opencv

Projects

- **FADE** Article, 2020

  Developing of an extensible and efficient **Spark** framework for alignment-free genomic analysis. Monte Carlo simulations on the output of AF functions can seamlessly be afforded for either small or huge collections of sequences. An experimental study has been conducted on a 25 nodes cluster of the *Consortium GARR*. See [3] for details.

- **FASTdoopC** Article, 2020

  Developing two general methods that can be used to adapt standard FASTA/Q data compression programs to work as **Hadoop** splittable data compression codecs. The use of the presented methods require very little, or none at all, programming and knowledge of Hadoop to carry out a rather complex task. The experimental analysis confirmed both space and time savings for the storage and processing of FASTA/Q files. See [4] for details.

- **EV3 Autonomous Driving** LegoLab, 2018

  Developing a (naive) autonomous driving system by using the **Lego Mindstorm EV3** device, provided by the LegoLab laboratory. A smartphone was positioned upon the car to be used as camera sensor and on-board computer. An **Android** application has been implemented for the acquisition of the camera frames, the extraction of features by using **Computer Vision** techniques and the communication, via bluetooth, for the right trajectory to follow.

Activities and Seminars

- **TPC member** (HPCC-2020, virtual conference) 14/12/20-16/12/20
  Technical Program Committee (TPC) Member for IEEE HPCC-2020 (The 2020 IEEE International Conference on High Performance Computing and Communications).
- **Speaker at the DASC Conference** (DASC 2020, virtual conference) 17/08/2020
  I presented the paper High-Performance Computing with TeraStat, during the 18th IEEE International Conference on Dependable, Autonomic and Secure Computing.
- **Advanced School in AI** (ISTC-CNR, Rome, Italy) 10/19 - 04/20
  I attended the courses (https://as-ai.org) offered by the ISTC institute of the CNR in order to improve my skills in the field of Artificial Intelligence.
- **Research project member** (Dept. Statistical Science, Sapienza University of Rome) 22/02/2019
  - *role: Investigator*
  - *scope: Progetto di Ricerca di Ateneo, Sapienza University of Rome*
  - *person in charge: Umberto Ferraro Petrillo*
  - *protocol number: RM11816435FE1EA2*
- **Speaker at the AIBH Workshop** (IEEE BIBM 2018, Madrid, Spain) 03/12/2018
  I presented the paper Partial Parallel Error Correction of Pair-end Short Reads using GPU, during the Artificial Intelligence Techniques for Biomedicine and Healthcare Workshop.

- **Seminar on R packages** (Dept. Statistical Science, Sapienza University of Rome) 22/09/2018
  During the course Laboratorio di Software Statistici, taught by prof. Stefania Gubbiotti, I held a lesson in which were explained the steps required for the creation of a software package in R.
- **Laboratory on Neo4j Database** (Dept. Statistical Science, Sapienza University of Rome) 03/05/2018
  I held a laboratory during the course Gestione di Big Data, taught by prof. Umberto Ferraro Petrillo, where the students practiced with the syntax and concepts of Neo4j query language on a real-world use case.

Publications

[1] E. Bompiani, U. F. Petrillo, G. J. Lasinio, and **F. Palini**, "High-performance computing with TeraStat", in *2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing (DASC)*, IEEE, 2020.

[2] C. Guerra, S. Joshi, Y. Lu, **F. Palini**, U. F. Petrillo, and J. Rossignac, "Rank-similarity measures for comparing gene prioritizations: A case study in autism", *Journal of Computational Biology*, 2020.

[3] U. F. Petrillo, **F. Palini**, G. Cattaneo, and R. Giancarlo, *Alignment-free Genomic Analysis via a Big Data Spark Platform*, 2020. arXiv: 2005.00942 [cs.DC].

[4] U. F. Petrillo, **F. Palini**, G. Cattaneo, and R. Giancarlo, *FASTA/Q Data Compressors for MapReduce-Hadoop Genomics:Space and Time Savings Made Easy*, 2020. arXiv: 2007.13673[cs.DC].

[5] G. Cattaneo, U. F. Petrillo, R. Giancarlo, and **F. Palini**, "A Spark algorithmic paradigm for spaced words alignment-free classification, with focus on phylogeny", in *BITS2019 Bioinformatics Italian Society Annual Meeting*, Bioinformatics Italian Society, 2019.

[6] U. F. Petrillo and **F. Palini**, "Partial parallel error correction of pair-end short reads using GPU", in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, 2018, pp. 2114–2116.

Autorizzo la pubblicazione ai sensi della normativa sulla trasparenza (D.lgs. 33/2013).