

Claudio Scalzo

SOFTWARE ENGINEER INTERN AT SAP • SOFTWARE ENGINEERING AND DATA SCIENCE MSc STUDENT

☎ (+39) 346 24 55 116 • ✉ claudio.scalzo@outlook.com • 🌐 github.com/claudioscalzo • 🔗 linkedin.com/in/claudioscalzo

Born in Italy in 1994. Passionate about software development in the cloud computing landscape. Strong analytical thinking and team-working skills. Successful results due to the passion for this field and the high precision in projects and tasks fulfillment. Languages: Italian (MT), English (C1), French (intermediate).



Education

Master of Science in Data Science and Engineering

EURECOM / TELECOM PARISTECH

- Double Degree program between Telecom ParisTech and Politecnico di Torino

Sophia Antipolis, France

Sep. 2017 - Mar. 2019

Master of Science in Software Engineering

POLITECNICO DI TORINO

Turin, Italy

Sep. 2016 - Mar. 2019

Bachelor's Degree in Software Engineering

POLITECNICO DI TORINO

Turin, Italy

Sep. 2013 - Jul. 2016

- with the highest grade, 110/110

Experience

SAP France

Paris, France

SOFTWARE ENGINEER / DATA ENGINEER INTERN

Jul. 1, 2018 - Dec. 31, 2018

- Worked on a Data Lake platform based on Microsoft Azure, adding a framework to extend the post-processing phase with machine learning capabilities, and the pre-processing phase with external data ingestion and ADLS data transfer.
- Leveraged Spark parallelization, exploiting the distributed computing power offered by the HDInsight clusters, storing raw/processed data on the ADLS, and making the output available in Hive tables consumable by the business data warehouse (on SAP HANA).
- Managed Jenkins continuous delivery pipelines for both testing and production environments; provided support, troubleshooting and solution proposals during the daily team operations.

Skills

Languages Python • C • Java • SQL (Oracle PL/SQL, HiveQL) • JavaScript • SPARQL • MATLAB • Bash / AWK / Sed

Big Data Apache Spark • Apache Hive • MapReduce & HDFS • NoSQL Architectures

Cloud Computing Microsoft Azure (HDInsight, ADLS) • Apache Ambari cluster monitoring

Machine Learning Spark ML • scikit-learn • Keras • Deep Learning (NNs, CNNs, RNNs) • Probabilistic Machine Learning • NLP

VCS, CI, CD & Others Git • Jenkins CI/CD • Jupyter / Zeppelin Notebooks • Dialogflow (+ BotKit)

Projects

Virtual Assistant for answering music related questions

2018 | [GITHUB.COM/D2KLAB/MUSIC-CHATBOT](https://github.com/D2KLAB/MUSIC-CHATBOT) • [CHATBOT.DOREMUS.ORG](https://chatbot.doremus.org)

- Virtual assistant development (in the chatbot and vocal assistant forms), capable of answering music related questions and providing detailed graphical results. Informations extracted from the *DOREMUS* knowledge base, queried using the SPARQL language.
- Built with Node.js, using the BotKit framework. Trained Google Dialogflow as NLP. Facebook Messenger, Slack and Google Assistant support.
- Contributed with two accepted pull requests to the `botkit-middleware-dialogflow` author, for concurrency and language support.

Team leader for an optimization project for TIM / SWARM Joint Open Lab

2016 | [GITHUB.COM/CLAUDIOSCALZO/COIOTE](https://github.com/claudioscalzo/COIOTE)

- Solving of a *VRP (Vehicle Routing Problem)* optimization problem proposed by *TIM* and the *SWARM Joint Open Lab*, for an IOT project named *ColoTe*. Multistart tabu-search approach, written in Java (with the *OpenTS* Java library).
- Achieved 1st position in the final ranking. Secured great comprehension of metaheuristics. Improved algorithmic and team-working skills.

Music Recommendation System with Spark MLlib and with a custom approach

2018 | [GITHUB.COM/LOMLUCA/AML/TREE/MASTER/REC-SYS](https://github.com/LOMLUCA/AML/TREE/MASTER/REC-SYS)

- Development of a music recommendation system, starting from a *Last.fm* dataset containing user IDs, artist IDs and the related play counts.
- Written in Python, following and comparing two different paths: the Spark MLlib *ALS (Alternating Least Squares)* algorithm and a custom algorithm which, for this task, outperformed the *ALS* performances.