

# Daniele Foroni | Ph.D.

✉ [daniele.foroni@gmail.com](mailto:daniele.foroni@gmail.com) • [forons.github.io](https://forons.github.io) • [twitter forons](https://twitter.com/forons) • [forons](https://forons.com) • [in danieleforoni](https://www.linkedin.com/in/danieleforoni)

Senior Data Engineer and Senior Data Scientist with 5+ years of experience in designing, scaling, and deploying data processing pipelines in production, as well as in creating models and optimization algorithms in the Smart Cities context. Most recently developed an optimization algorithm for traffic light management that reduces of up to 20 % the driving time.

## Work Experience

- From 2019 **Senior Big Data Research Engineer**, *Huawei MRC (Munich Research Center)*, Munich, Germany.  
Aug–Now Research and develop Big Data technologies with special focus on real-time and streaming technologies, and techniques for traffic management optimization in smart city environments. I create algorithms for traffic optimization, which reduced up to 20 % the driver delay, and for anomaly discovery in traffic patterns. Moreover, I am in charge for delivering to product the features developed by the whole team. I am also the codebase, CI, and server maintainer.
- From 2018 **CTO and Full-Stack Developer**, *IOHub*, Rovereto, Italy.  
Oct–Now Start-up that targets privacy-aware solutions for both citizens and businesses about the Italian electronic invoicing system. CEO: [Gianmaria Conti](#)
- 2017–2018 **Research Engineer Intern**, *Huawei ERC*, Munich, Germany.  
Visiting researcher working on dynamic rescheduling resources for streaming queries in a distributed environment based on the goal metrics defined by the user. Advisors: [Stefano Bortoli](#) and [Radu Tudoran](#)
- 2016–2017 **Data Scientist**, *University of Trento and Saidea SRL*, Trento, Italy.  
Development of XplainBox, an industrial tool that reads and stores data in a streaming fashion, and on it performs analyses and applies ML algorithms.
- 2014 **Data Engineer Intern**, *Engineering Tributi SPA*, Trento, Italy.  
Feb–Nov Development of an entity resolution tool that, with a set of rules, compares similar entities, avoiding meaningless comparisons. Advisor: [Nicola Mezzetti](#)

## Education

- 2014–2019 **Ph.D. in Computer Science**, *University of Trento, Italy, Program funded by SkillLab - TIM (Telecom Italia)*.  
Thesis: *Putting Data Quality in Context: How to generate more accurate analyses*. Advisor: [Yannis Velegarakis](#)
- 2012–2014 **Master of Science in Computer Science**, *University of Trento, Italy*.
- 2009–2012 **Bachelor of Science in Computer Science**, *University of Trento, Italy*.

## Skills

- Programming Python · Java · Scala · C# · SQL (Postgres, MySQL) · Javascript · PHP
- Big Data Apache Spark · Apache Flink · Apache Kafka · Apache Hadoop · NoSQL
- Certifications [Scalable Machine Learning with Apache Spark](#) · [Python Bootcamp for Data Science 2021](#)
- Tools & OS Bash · Git · QGIS · Linux · Mac OS · Windows
- AI Machine Learning · Data Mining

## Teaching, Lectures, and Talks

- 2021 **Lecturer, Apache Spark - The Art of Distributing**, *University of Trento, Trento, Italy*.  
Lecture and lab sessions on Apache Hadoop and Spark for the *Big Data* Master course. Reference: [Yannis Velegarakis](#)
- 2021 **Invited Talk, Apache Spark - Notions and Lab Session**, *Utrecht University, Utrecht, The Netherlands*.  
Talk and lab sessions on Apache Hadoop and Spark for the *Data Intensive* Master course. Reference: [Yannis Velegarakis](#)
- 2020 **Invited Talk, From Hadoop to Spark - A Big Data Journey**, *University of Trento, Trento, Italy*.  
Talk about Big Data tools, with focus on Apache Hadoop and Spark with its libraries. Reference: [Yannis Velegarakis](#)
- 2015–2019 **Teaching Assistant, Big Data and Social Networks course**, *University of Trento, Trento, Italy*.  
Master course on Big Data storage (NoSQL, HDFS) and analysis (MapReduce, Spark, ML models, Streaming, Graph processing), with Data Mining notions. Professor: [Yannis Velegarakis](#)

- 2017 **Lecturer, Big Data for Data Scientists, Telecom Italia - TIM, Trento, Italy.**  
Specialized course for data scientists in TIM on HDFS, Spark, ML, and Streaming. Reference: [Michele Vescovi](#)
- 2015–2016 **Teaching Assistant, Programming for Mathematicians course, University of Trento, Trento, Italy.**  
Bachelor course for introducing Math students to programming. Professor: [Roberto Zunino](#)

## Selected Publications

- 2021 *Estimating the Extent of the Effects of Data Quality through Observations* **ICDE 2021**  
D. Foroni, M. Lissandrini, and Y. Velegrakis [doi:10.1109/ICDE51399.2021.00176](#)
- 2021 *The F4U System for Understanding the Effects of Data Quality* **ICDE 2021**  
D. Foroni, M. Lissandrini, and Y. Velegrakis [doi:10.1109/ICDE51399.2021.00312](#)
- 2021 *OBELISC: Oscillator-Based Modelling and Control using Efficient Neural Learning for Intelligent Road Traffic Signal Calculation* **ECML-PKDD 2021**  
C. Axenie, D. Foroni, A. Wieder, M. Al Hajj Hassan, P. Sottovia, M. Grossi, . . . , S. Bortoli, and G. Brasche [doi:10.1007/978-3-030-86514-6\\_27](#)
- 2021 *TRAMESINO: Traffic Memory System for Intelligent Optimization of Road Traffic Control* **AALTD 2021**  
C. Axenie, D. Foroni, A. Wieder, M. Al Hajj Hassan, P. Sottovia, M. Grossi, . . . , S. Bortoli, and G. Brasche [Av. upon request](#)
- 2020 *Real-time Traffic Jam Detection and Congestion Reduction Using Streaming Graph Analytics* **BSD 2020**  
Z. Abbas, P. Sottovia, M. Al Hajj Hassan, D. Foroni, S. Bortoli [doi:10.1109/BigData50022.2020.9378068](#)
- 2018 *Moirá: A Goal-Oriented Incremental Machine Learning Approach to Dynamic Resource Cost Estimation in Distributed Stream Processing Systems* **BIRTE 2018**  
D. Foroni, C. Axenie, S. Bortoli, M. Al Hajj Hassan, . . . , and Y. Velegrakis [doi:10.1145/3242153.3242160](#)
- 2018 *STARLORD: Sliding Window Temporal Accumulate-Retract Learning for Online Reasoning on Datastreams* **ICMLA 2018**  
C. Axenie, R. Tudoran, S. Bortoli, M. Al Hajj Hassan, D. Foroni, and G. Brasche [doi:10.1109/ICMLA.2018.00181](#)
- 2018 *Data Management and Smart Cities* **IEEE Smart Cities 2018**  
M. Brugnara, C. Consonni, D. Foroni, S. Pasupathipillai, G. Preti, P. Sottovia, and Y. Velegrakis [White Paper](#)

## Projects

- 2019–Now **Data Quality Measurement, Ongoing work.**  
Built the theoretical framework for assessing the dataset quality for the application (task) the user has in mind. This approach is being applied by big companies and additionally it has been extensively tested with multiple tasks and datasets.
- 2020–Now **Entity Matching through data profiling, Ongoing work.**  
Exploited the intrinsic value of the data through a profile for improving the accuracy of the duplicates found.
- 2017 **Data Lake analysis: SQL, statistics, Data Mining, and ML techniques, Industrial project.**  
Design a flow of Data Mining and Machine Learning algorithms, combined with statistics and queries defined by the users, to show the key insights from the data.

## Awards and Leadership

- 2021 **Huawei Year Award.** I got the yearly award for the exceptional work in the previous year.
- 2018–Now **Student guidance.** I guided Master and Ph.D. students during their internships.
- 2019 **Soccer Data Challenge.** 1<sup>st</sup> place at the SoBigData Challenge about soccer Match Analysis.
- 2015 **Travel Awards.** Databricks grant based on the research proposal and achievement.

## Other activities

- PC member EDBT 2021 · SEA Data 2021 · SEA Data 2020
- Reviewer VLDB · ICDE · KDD · SIGMOD Record · IEEE BigData · EDBT · ECML-PKDD · JPDC · ICMLA · CSGRID

## Languages

**English:** Full professional proficiency (C1) · **German:** Basic Knowledge (A1) · **Italian:** Native