Matteo **Brucato**

Researcher · Data Systems Group · Microsoft Research

Microsoft Research, Redmond WA, USA

mbrucato@microsoft.com | https://matteo-brucato.github.io/ | MatteoBrucato

Summary

Researcher at Microsoft Research, in the Data Systems Group.

Ph.D. graduate from the University of Massachusetts Amherst with a focus on prescriptive analytics and database systems. My doctoral thesis, titled "Package Queries: Enabling Declarative and Scalable Prescriptive Analytics in Relational Data," introduced new methods to implement prescriptive analytics in relational databases through declarative queries. This research contributes to simplifying the complexity of analytics applications and enhancing scalability. With a robust background in the theoretical and practical aspects of database management and analytics, I am equipped to deliver valuable insights and innovative solutions in technology and analytics environments.

Research Interests

Core areas:
Data Management Systems: Enhancing capabilities to support predictive and prescriptive analytics. **Scalable Machine Learning**: Developing methods and systems that improve fairness and efficiency in machine learning applications. **Decision Support Systems**: Building advanced tools to facilitate complex decision-making processes in uncertain environments.

Selected topics:
Prescriptive Analytics: Extending query engines to support in-database prescriptive analytics. **Markov Decision Processes**: Creating scalable solutions for sequential decision-making problems. **Fair Machine Learning**: Integrating fairness into existing and new machine learning frameworks. **Personalized Document Summarization**: Designing example-driven summarization techniques. **Temporal Information Retrieval**: Advancing the retrieval of time-based data for enhanced contextual understanding.

Professional Experience

Industry

Microsoft Research — *Redmond, WA, USA*
Researcher — *Oct 2021 - present*

Microsoft Research — *Redmond, WA, USA*
Research Intern — *May - Aug 2017*
Worked under the supervision of **Kaushik Chakrabarti** on natural language to SQL using neural networks.

IBM Watson — *T. J. Watson Research Center, Yorktown Heights, NY, USA*
Research Intern, Watson Health — *Jun - Aug 2016*
Worked under the supervision of **Shilpa Mahatma** and **Yajuan Wang** on "*conceptual queries*" for healthcare applications.

CINECA — *Casalecchio di Reno, Bologna, Italy*
Software Engineering Intern — *Jun - Oct 2012*
- Responsible for developing and managing a data warehouse and analytics system for a nationwide videoconferencing system.
- Duties included generating aggregated data and analytics, deployed through usable interfaces to let decision makers understand usage and performance of the videoconferencing system.

Maggiore Hospital — *Bologna, Italy*
Software Developer (volunteer) — *2011 - 2013*
- Co-developed software to help the Emergency Room visit their patients using portable ultrasound technology.
- Software deployed on medical tablets to help doctors collect patient's information on the field.
- The collected data was then used to make better informed decisions in the Emergency Room.

Academia

DREAM Lab, University of Massachusetts

Amherst, MA, USA

Research Assistant — *Sep 2013 - May 2021*

- Worked with **Prof Peter J. Haas**, **Prof Alexandra Meliou**, and **Prof Azza Abouzied** (NYU Abu Dhabi) to create **Package Builder**, a system that extends query engines to support the generation of *packages*, i.e., collections of tuples with global properties.
- Technical paper at VLDB 2024
- Technical paper at SIGMOD 2020
- Demo papers at VLDB 2020
- Invited journal paper at CACM Research Highlights 2019.
- Invited journal paper at the VLDB Journal 2018.
- Invited journal paper at SIGMOD Records 2017.
- Technical paper at VLDB 2016; workshop paper and demo paper at VLDB 2014; undergraduate research poster at SIGMOD 2014.
- Project website: packagebuilder.cs.umass.edu.
- DREAM Lab website: dbgroup.cs.umass.edu.
- Code: Scalable-PaQL-Queries on GitHub

Design Technology Lab, NYU Abu Dhabi

Abu Dhabi, UAE

Visiting Ph.D. Candidate — *Mar - May 2016*

- Paper: Redistributing Funds across Charitable Crowdfunding Campaigns (arXiv:1706.00070).

AMPLab, UC Berkeley

Berkeley, CA, USA

Visiting Ph.D. Student — *May - Aug 2015*

- Worked on extensions to Package Builder to identify ideal packages by example, under NYU Summer internship.

Data-Intensive Systems Lab, Århus University

Århus, Denmark

Visiting M.S. Student — *Jan - Jun 2013*

- Worked on M.S. thesis on *Temporal Information Retrieval* under the supervision of **Prof Christian S. Jensen**.
- Actively collaborated with researchers within the group.
- Published two papers in main international conferences in the fields of NLP and IR.

Database Lab, UC Riverside

Riverside, CA, USA

Undergraduate Research Intern (volunteer) — *Jan - Jun 2013*

- Worked closely with a Ph.D. candidate and **Prof Vassilis J. Tsotras** on query optimization techniques for query result diversification.
- Developed a query parser for a SQL-like diversification query language.
- Devised and developed a query optimizer to select optimal diversification query plans.
- Developed client/server technology to issue and evaluate diversification queries.

Honors & Awards

2020	**Best Demo Award**, VLDB 2020	*Tokyo, Japan*
2020	**Best Demo Runner-up Award**, VLDB 2020	*Tokyo, Japan*
2019	**Krithi Ramamritham Computer Science Scholarship**	*Amherst, MA, USA*
2018	**CACM Research Highlight Award (Invited journal paper in CACM Research Highlights)**	*Amherst, MA, USA*
2017	**ACM SIGMOD Research Highlight Award (invited journal paper in SIGMOD Records)**	*Amherst, MA, USA*
2017	**Invited journal paper in the VLDB Journal, Special Issue on Best Papers of VLDB 2016**	*Amherst, MA, USA*
2016	**Best Papers of**, VLDB 2016	*New Delhi, India*
2013	**M.S. with Honors**, University of Bologna	*Bologna, Italy*
2013	**Best Young Researcher Award Nominee**, RANLP 2013	*Hissar, Bulgaria*
2013	**Merit Scholarship**, awarded to 0.2% of all students of the University of Bologna	*Bologna, Italy*
2012	**Research Scholarship**, University of Bologna	*Bologna, Italy*
2012	**Merit Certificate**, University of Bologna	*Bologna, Italy*
2012	**Complimentary AAAI Membership**, Stanford Online AI Course	*Stanford, CA, USA*
2011	**B.S. with Honors**, University of Bologna	*Bologna, Italy*
2011	**Merit Certificate**, University of Bologna	*Bologna, Italy*
2010	**Merit Certificate**, University of Bologna	*Bologna, Italy*
2010	**Exchange Scholarship**, University of Bologna	*Bologna, Italy*

Publications

1. **Matteo Brucato**, Tarique Siddiqui, Wentao Wu, Vivek Narasayya, Surajit Chaudhuri
   Wred: Workload Reduction for Scalable Index Tuning
   SIGMOD 2024 *(conference) pp.1-26*
2. Anh L. Mai, Pengyu Wang, Azza Abouzied, **Matteo Brucato**, Peter J. Haas, Alexandra Meliou
   Scaling Package Queries to a Billion Tuples via Hierarchical Partitioning and Customized Optimization
   VLDB 2024 *(conference) pp.1146-1158*
3. S. G. Rizzo, **M. Brucato**, D. Montesi
   Ranking Models for the Temporal Dimension of Text

ACM Transactions on Information Systems (TOIS) 2023 *(journal) pp.1-34*

4. N. Yadav, **M. Brucato**, A. Fariha, O. Youngquist, J. Killingback, A. Meliou, P. J. Haas
SubSumE: A Dataset for Subjective Summary Extraction from Wikipedia Documents
NewSum @ EMNLP 2021 *(conference)*
5. S. B. Nashed, J. Svegliato, **M. Brucato**, C. Basich, R. Grupen, S. Zilberstein
Solving Markov Decision Processes with Partial State Abstractions
ICRA 2021 *(conference) pp.813-819*
6. **M. Brucato**, M. Mannino, A. Abouzied, P. J. Haas, A. Meliou
sPaQLTooLs: A Stochastic Package Query Interface for Scalable Constrained Optimization
VLDB 2020 *(conference) pp.2881-2884*
Best Demo Award. Video: `https://www.youtube.com/watch?v=1v5KevGKbfE`
7. A. Fariha, **M. Brucato**, P. J. Haas, A. Meliou
SuDocu: Summarizing Documents by Example
VLDB 2020 *(conference) pp.2861-2864*
Best Demo Runner-Up Award. Video: `https://www.youtube.com/watch?v=ZGzF6cBSYjo`
8. **M. Brucato**, N. Yadav, A. Abouzied, P. J. Haas, A. Meliou
Stochastic Package Queries in Probabilistic Databases
SIGMOD 2020 *(conference) pp.269-283*
9. **M. Brucato**, A. Abouzied, A. Meliou
Scalable Computation of High-Order Optimization Queries
Communications of the ACM (CACM) 2019 *(journal) pp.108-116*
10. **M. Brucato**, A. Abouzied, A. Meliou
Package queries: Efficient and scalable computation of high-order constraints
The VLDB Journal 2018 *(**Special Issue on Best Papers of VLDB 2016**) (journal) pp.693-718*
11. **M. Brucato**, A. Abouzied, A. Meliou
A Scalable Execution Engine for Package Queries
SIGMOD Record 2017 *(**ACM SIGMOD Research Highlight Award 2017**) (journal) pp.24-31*
12. **M. Brucato**, A. Abouzied, C. Blauvelt
Redistributing Funds across Charitable Crowdfunding Campaigns
(tech report) arXiv:1706.00070 2017
13. **M. Brucato**, J. F. Beltran, A. Abouzied, A. Meliou
Scalable Package Queries in Relational Database Systems
VLDB 2016 *(conference) pp.576-587*
Best papers of VLDB 2016
Invited journal paper in the VLDB Journal (Special Issue on Best Papers of VLDB 2016)
ACM SIGMOD Research Highlight Award (Invited journal paper in SIGMOD Records 2017)
CACM Research Highlight Award (Invited journal paper in CACM Research Highlight 2018)
14. **M. Brucato**, A. Abouzied, and A. Meliou
Improving Package Recommendations through Query Relaxation
Data4U @ VLDB 2014 *(conference) pp.13-18*
15. **M. Brucato**, R. Ramakrishna, A. Abouzied, A. Meliou
PackageBuilder: From Tuples to Packages
VLDB 2014 *(conference) pp.1593-1596*
16. K. Fernandes, **M. Brucato**, R. Ramakrishna, A. Abouzied, A. Meliou
PackageBuilder: Querying for packages of Tuples
SIGMOD 2014 *(conference) pp.1613-1614*
17. **M. Brucato**, D. Montesi
Metric Spaces for Temporal Information Retrieval
ECIR 2014 *(conference) pp.385-397*
18. **M. Brucato**, L. Derczynski, H. Llorens, K. Bontcheva, C. S. Jensen
Recognising and Interpreting Named Temporal Expressions
RANLP 2013 *(conference) pp.113-121*
Best Young Researcher Award Nominee

Education

UMASS AMHERST

Amherst, MA, USA

PH.D. IN COMPUTER SCIENCE

2013 - 2021

- Co-advisors: Peter J. Haas, Alexandra Meliou, and Azza Abouzied (NYU)
- Thesis: *Package Queries: Enabling Declarative and Scalable Prescriptive Analytics in Relational Data*

UNIVERSITY OF BOLOGNA *Bologna, Italy*

M.S. in Computer Science *2011 - 2013*

- Summa cum laude
- Thesis Advisor: Danilo Montesi
- Thesis: *Temporal Information Retrieval*

UNIVERSITY OF BOLOGNA *Bologna, Italy*

B.S. in Computer Science *2008 - 2011*

- Summa cum laude
- Thesis Advisor: Danilo Montesi
- Thesis: *Mobile systems for patient-reported outcomes*

Exchange Programs

ÅRHUS UNIVERSITY *Århus, Denmark*

Exchange Program during M.S., Computer Science *Jan - Jun 2013*

- Advisor: Christian S. Jensen

UC RIVERSIDE *Riverside, CA, USA*

Exchange Program during B.S., Computer Science *Sep - Jun 2011*

- Advisor: Vassilis J. Tsotras

Advising & Mentoring

Research Advisor for Ph.D. students Riddho Ridwanul and Anh Mai *Amherst, MA*

UMass Amherst *2021*

Riddho (UMass Amherst) and Anh (NYU Abu Dhabi) are first-year Ph.D. students and working on package queries.

Research Advisor for M.S. students Oscar Youngquist and Julian killingback *Amherst, MA*

UMass Amherst *2021*

Oscar is working on enhanching SuDocu, our system for extractive document summarization by example. Julian helped us collecting user-annotade data from Mechanical Turk.

Research Advisor for undergraduate student Agam Sandhu *Amherst, MA*

UMass Amherst *2020*

Agam Sandhu worked on scalable stochastic constrained optimization under costly sampling methods.

Research Advisor for PhD student Ahmad Chatha (NYU) *Amherst, MA*

UMass Amherst *2020*

Ahmad Chatha worked on combining scalable constrained optimization methods with scenario-based methods for stochastic optimization.

Project Mentor (in collaboration with Microsoft Research) *Amherst, MA*

UMass Amherst *2020*

Mentored MS students (Abdul Hannan Kanji, Parul Gupta, Ge Gao, Xinlu Liu) for their Independent Class project (CS 696DS), in collaboration with Microsoft Research, to improve the Fairlearn library (`https://fairlearn.github.io`).

Research Advisor for MS student Kanchi Masalia *Amherst, MA*

UMass Amherst *2020*

Kanchi Masalia worked on parts of SuDocu, our system for personalized efficient document summarization by example.

Research Advisor for undergrad Genglin Liu *Amherst, MA*

UMass Amherst *2020*

Genglin Liu worked on the frontend of SuDocu, our system for personalized efficient document summarization by example.

Research Mentor for undergraduate student Kyaw Htet *Amherst, MA*

UMass Amherst *Fall 2019*

Kyaw worked on analyzing scenario reductions methods for stochastic optimization.

Project Mentor (in collaboration with Microsoft) *Amherst, MA*

UMass Amherst *Spring 2019*

Mentored a group of four MS students (Katie House, Shanu Vashishtha, Deeksha Razdan, Cheng-Yin Eng) for their Independent Class project (CS 696DS), in collaboration with Microsoft to produce notebooks that showcase their Machine-as-a-service platform Azure ML.

Research Mentor for undergraduate student Nishad Ranade *Amherst, MA*

UMass Amherst *2018–2019*

Nishad worked on Monte Carlo scenario reductions methods.

Research Mentor for 1st-year PhD student Nishant Yadav *Amherst, MA*

UMass Amherst *2018*

Nishant and I worked together on stochastic package queries.

Research Mentor for graduate student Sandhya Sankaranarayanan *Amherst, MA*

UMass Amherst *Fall 2015*

Sandhya and I worked together on extending the package query language PaQL.

Research Mentor for undergraduate student Kevin Fernandez *Amherst, MA*

UMass Amherst *Spring 2014*

Kevin and I worked together on evaluation techniques for package queries. This work was invited to present at the *ACM SIGMOD 2014 Undergraduate Research Poster Competition*.

Research Mentor for graduate student Rahul Ramakrishna *Amherst, MA*

UMass Amherst *2013 - 2014*

Rahul and I worked together on interface abstractions for package queries. Our work was demonstrated at *VLDB 2014*.

Teaching

Teaching Assistant for Graduate Database Systems (CS 645) *Amherst, MA*

UMass Amherst *Spring 2021*

Office hours, question answering, grading, projects.

Teaching Assistant for Undergraduate Database Systems (CS 345) *Amherst, MA*

UMass Amherst *Fall 2020*

Office hours, question answering, grading and labs.

Teaching Assistant for Undegraduate Class "A Networked World" (INFO 203) *Amherst, MA*

UMass Amherst *Spring 2018*

Homework grading.

Teaching Assistant for Undergraduate Introduction To Computation (CS 250) *Amherst, MA*

UMass Amherst *Fall 2017*

Responsible for discussion sessions, occasional lectures, office hours, homework and discussion grading.

Teaching Assistant for Undergraduate Database Systems (CS 345) *Amherst, MA*

UMass Amherst *Fall 2015*

Responsible for occasional lectures, project advising, office hours, homework and project grading.

Professional Service & Leadership

Program Committee

- *SIGMOD 2025*
- *VLDB 2025*
- *DataPlat workshop @ ICDE 2024*
- *IEEE Transactions of Knowledge and Data Engineering (TKDE) 2020*

Guest Editorial Board

- *Information Processing and Management*, special issue on *Time and Information Retrieval*, Elsevier, 2014/2015.

External Review Committee

- VLDB 2017
- SIGMOD 2015
- VLDB 2015
- RANLP 2013

Lab Seminar Organizer *Amherst, MA, USA*

UMass Amherst Database Lab *2014 - 2015*

- Coordinated weekly discussion seminar, including invited talks and paper readings, for the UMass Amherst Database Research group (∼10 Ph.D. students and faculty), and webmaster duties.

Student Seminar Leader *Amherst, MA, USA*

UMass Amherst Database Lab *Fall 2015*

- Organized and coordinated a weekly discussion seminar among Ph.D. and M.S. students in the UMass Amherst Database Research group to foster research collaborations, exchange of ideas, group discussions, and group cohesion.

Talks

Industry

Package Queries: Enabling Declarative and Scalable Prescriptive Analytics in Relational Data *Virtually*

Megagon Labs *Jan 2021*

Stochastic Package Queries for In-Database Prescriptive Analytics *Virtually*

Sisu Data *Sep 2021*

Scalable Package Queries in Relational Database Systems *Redmond, WA, USA*
Microsoft Research *Jun 2017*

Conceptual Queries for Multiple Data Sources *Austin, TX, USA*
IBM Watson *Aug 2016*

Package Queries: From Tuples to Packages *Berkeley, CA, USA*
Berkeley Labs *Jul 2015*

Academia

Package Queries: Scalable Prescriptive Analytics Close to the Data *Virtually*
Georgia Tech *Mar 2021*

Package Queries: Scalable Prescriptive Analytics Close to the Data *Virtually*
Carnegie Mellon University *Mar 2021*

Package Queries: Enabling Declarative and Scalable Prescriptive Analytics in Relational Data *Virtually*
Harvard University, DASlab *Jan 2021*

Stochastic Package Queries For In-database Optimization Under Uncertainty *Virtually at INFORMS 2020*
INFORMS 2020 *Nov 2020*

sPaQLTooLs: A Stochastic Package Query Interface for Scalable Constrained Optimization *Tokyo, Japan (virtually)*
VLDB 2020 *Sep 2020*
- Video: https://www.youtube.com/watch?v=1v5KevGKbfE

Stochastic Package Queries in Probabilistic Databases *Portland, OR, USA (virtually)*
SIGMOD 2020 *Jun 2020*
- Video: https://www.youtube.com/watch?v=UrEOEplK06I

Package Queries: Enabling Declarative and Scalable Prescriptive Analytics in Relational Data *Cambridge, MA, USA*
MIT CSAIL, Data Systems Group *Feb 2020*

Stochastic Package Queries for In-Database Constrained Optimization Under Uncertainty *Cambridge, MA, USA*
MIT, North East Database Day, Samberg Conference Center *Jan 2020*

Scalable Package Queries in Relational Database Systems *New Delhi, India*
VLDB 2016 *Sep 2016*
- Slides: packagebuilder.cs.umass.edu/papers/scalable-paql-presentation-share.pdf

Package Queries: From Tuples to Packages *Berkeley, CA, USA*
AMPLab, UC Berkeley *Jun 2015*

Improving Package Recommendations through Query Relaxation *Hangzhou, China*
VLDB 2014 *Sep 2014*
- Slides: packagebuilder.cs.umass.edu/papers/package-recomm-slides-web.pdf

Metric Spaces for Temporal Information Retrieval *Amsterdam, the Netherlands*
ECIR 2014 *Apr 2014*

Metric Spaces for Temporal Information Retrieval *Bologna, Italy*
University of Bologna *Dec 2013*

Metric Spaces for Temporal Information Retrieval *Århus, Denmark*
Århus University *Jan 2013*