Venkata Vamsikrishna Meduri

CONTACT INFORMATION

555 Bailey Ave
San Jose, CA 95141

Webpage: `https://vamsikrishna1902.github.io`
E-mail: Vamsi (dot) Meduri (at) ibm (dot) com

RESEARCH INTERESTS

Hybrid Search and Generative AI on Vector Databases,
Query Processing and Optimization in Lakehouse Systems,
Data Integration, Information Retrieval

CURRENT EMPLOYMENT

IBM Research - Silicon Valley, San Jose, California

- Staff Research Scientist (Dec 2025 - Present)

IBM Research - Almaden, San Jose, California

- Staff Research Scientist (April 2024 - Dec 2025)
- Research Scientist (May 2022 - April 2024)

EDUCATION

Arizona State University, Tempe, Arizona, USA

Ph.D., Computer Science (Jan 2015-May 2022)

- *Dissertation Title*: "Human-in-the-Loop Machine Learning Systems for Data Integration and Predictive Analytics"
- *GPA*: 3.9 / 4.0
- *Advisor*: Prof. Mohamed Sarwat

National University of Singapore, Singapore

M.S., Computer Science (Jan 2008-Dec 2010)

- *Thesis Title*: "Exhaustive Reuse of Subquery Plans to stretch Iterative Dynamic Programming for Complex Query Optimization"
- *GPA*: 4.06 / 5.0
- *Advisor*: Prof. Kian-Lee Tan

Vellore Institute of Technology, Vellore, India

B.Tech., Computer Science, (2003-2007)

- *Undergrad Thesis Title*: "Query Optimization in Multidatabase Systems"
- *GPA*: 9.26 / 10.0
- *Advisor*: Prof. Ranichandra Dharmaraj

RESEARCH PROJECTS

Data-centric Generative AI (dcgenAI) for Vector Databases

Vector database capabilities were integrated into the IBM lakehouse using SQL table functions for embedding creation (IBM Slate) and vector search (Microsoft DiskANN). The TPC-DS schema was enhanced with external ontologies to improve embedding context and vector search recall. A semantic layer was built for the dcgenAI schema using materialized views to accelerate common aggregate queries. Evaluation demonstrated that hybrid search, combining vector and lexical search, is essential for achieving high recall on queries requiring exact matches.

PTO: A Workload-driven Predictive Table Optimizer for Lakehouse Systems

We designed and developed a predictive table optimizer (PTO) to automatically optimize Apache Iceberg table layouts for the Presto/Prestissimo (Velox) engine. To avoid exhaustively enumerating

and configuring the table optimization parameters, PTO leverages machine learning models and table sampling to estimate the skipping benefits of filter predicates in a given workload on a candidate space of data layouts to choose the most optimal data layout. Our experiments show that PTO reduces the average workload latency by 11% on TPC-H and 36% on TPC-DS benchmarks at SF 10K while speeding up scan-intensive, long latency queries by 3.4× and 11× respectively.

Delivery of Query Optimizer in Watsonx.data 2.0 and Order of Magnitude Performance Improvement for TPC-DS 100 TBytes on Fusion HCI

IBM research initiated and delivered enterprise grade query optimization in Watsonx.data. We initiated the idea of using the Db2 query optimizer as a disaggregated optimizer for complex Presto SQL queries, prototyped the initial proof-of-concept, and collaborated with Data & AI Business Unit to deliver the technology in Watsonx.data 2.0. We internally delivered better price performance compared to Databricks Photon engine on a query benchmark derived from public 100TB TPC-DS. We accomplished equal query runtime at less than 60% of the cost using Watsonx.data 2.0 with query optimizer and Presto C++ v0.286 on IBM Fusion HCI. In this project, I accomplished the following tasks.

- Evaluated the capabilities of Velox data cache and Presto fragment result cache (RaptorX) on TPC-H and TPC-DS workloads.
- For performance quantification of the IBM lakehouse, I evaluated Velox as a native worker runtime to quantify the data processing throughput and to meet the performance targets. I measured that Velox can perform close to IBM Db2 BLU.

Initiation and General Availability (GA) of IBM Data Lakehouse (Watsonx.data)

IBM Data Lakehouse became GA in July 2023. IBM Research (Almaden Research Center) initiated the effort for IBM to enter the growing data lakehouse market. Research closely worked with Data & AI Business Unit (Silicon Valley Lab, Toronto, India) in setting the strategy and delivering the product. IBM Data Lakehouse is built on open source PrestoDB enriched with IBM technologies to make it enterprise ready. IBM Data Lakehouse builds the foundation for Watsonx.data.

Active Learning for Ontology Mapping

We reconcile diverse ontologies into a unified ontology by performing schema mapping. In order to alleviate the need for large amounts of training data, we selectively query a human-in-the-loop via active learning to label informative concept pairs from the ontologies. We evaluate our proposed active learning strategies against state-of-the-art baselines.

Guided Data Analysis for Conversational Business Intelligence

We predict and recommend next query patterns to Business Intelligence (BI) analysts who issue Natural Language (NL) queries on the top of a conversational system. Here we use a high-level abstraction for both the queries and the underlying database schema. The schema is abstracted as a BI ontology graph that is used to create ontology-aware embeddings for NL queries and the corresponding next query search space representation.

AI for Human-Database Interaction

We develop large-scale systems applying concepts from the AI literature such as reinforcement learning (RL) to Human-Database Interaction. We compare state-of-the-art ML techniques for SQL query prediction during an interaction session against our adaptation of synthesis-based RNNs and Q-Learning.

A Unified Active Learning Framework for Entity Resolution

We build a unified, generalized active learning framework for entity matching that can mix-and-match different learners with applicable example selection algorithms. We compare various example

selectors and learners on a variety of quality and latency metrics.

Rule Discovery in Knowledge Bases: RuDiK

We discover rules for the positive and negative relationships between entity pairs in KBs while being robust to their errors and incompleteness. The rules are mined from the training examples by employing scalable, disk-based generation and validation techniques.

Program Synthesis for Entity Matching

We use a *program synthesis* solver named Sketch to generate concise and interpretable boolean expressions (rules) satisfying matching and non-matching assertions on the training data to perform entity matching.

Data Integration of the Electric System Schemata

We integrate real world schemata with a lot of inconsistencies and apply approximate entity matching and schema alignment techniques to reconcile electric system transmission, distribution and location data with diverse format. This project was a collaborative effort between the CASCADE team at ASU and Salt River Project (SRP) which is one of the primary electricity distributors in Arizona.

Prior Experience

IBM Almaden, San Jose, California

Summer Research Intern **May-August 2021**
"Active Learning for Ontology Mapping" in the Database group
Mentors: Abdul Quamar, Chuan Lei, Xiao Qin
Manager: Berthold Reinwald

IBM Almaden, San Jose, California

Summer Research Intern **June-September 2020**
"Conversational BI Recommendation" in the Database group
Mentors: Abdul Quamar, Chuan Lei, Vasilis Efthymiou
Manager: Fatma Özcan

IBM Almaden, San Jose, California

Summer Research Intern **May-August 2017 & 2018**
"Unified Active Learning" in the Scalable NLP group
Mentors: Lucian Popa, Min Li, Prithviraj Sen, *Manager*: Yunyao Li

SRM Research Institute, Bangalore, India

Member of the Research Group **April, 2013 - December, 2014**
Participated in the design and development of an E-learning portal, *Atiha*, and an energy management framework for a smart grid using applied machine learning.

Centrum Wiskunde en Informatica, Amsterdam, The Netherlands

Ph.D. Student **Jan, 2012 - Jan, 2013**
Developed parallel query processing algorithms for pre-clustered cache-conscious column store database on multi-core architecture

National University of Singapore, Singapore

Research Assistant **Dec, 2010 - Dec, 2011**
Design and development of database query processing algorithms for Phase Change Memory

Oracle India Private Limited, Bangalore, India

Associate Application Engineer **July, 2007 - December, 2007**

Developed Web applications in Java for Oracle 10g

Coding Skills Java, Python, C++, SQL, and some usage of Shell scripts

Coursework Statistical Machine Learning (Spring 2015), Semantic Web Mining (Spring 2015), Game Theory (Fall 2015), Database Management Systems Implementation (Spring 2016), Fundamentals of Statistical Learning (Fall 2016), Advanced Topics on Social Media Analysis (Fall 2017)

Teaching Assistant
CSE 511 - Data Processing at Scale (Spring 2021, Instructor)
CSE 412 - Database Management Systems (Spring 2017 & 2020)
CSE 310 - Data Structures and Algorithms (Fall 2019)
CSE 494 - Information Retrieval, Mining and Integration on the Internet (Spring 2015 & 2016)
CSE 110 - Principles of Programming in Java (Fall 2015)

Technical Accomplishments Berthold Reinwald, Hamid Pirahesh, Michael Kaufmann, Nasrullah Sheikh, Richard Sidle, Venkata Vamsikrishna Meduri, Zoltan Arnold Nagy, Ronald Barber, Pascal Spoerri, Gregory Kishi, Aditi Pandit, Ajay Gupta, Arin Mathew, Ashok Kumar, Austin Clifford, Calisto Zuzarte, Christian Zentgraf, Deepak Majeti, Ethan Zang, George Lapis, Jason Sizto, Sudheesh Kairali: **Delivery of Query Optimizer in Watsonx.data 2.0 and Order of Magnitude Performance Improvement for TPC-DS 100 TBytes on Fusion HCI**. IBM Research Accomplishment 2024.

Hamid Pirahesh, Berthold Reinwald, Larry Chiu, Ronald Barber, Richard Sidle, Scott Guthridge, Venkata Vamsikrishna Meduri, Nasrullah Sheikh, Frank Schmuck: **Initiation and GA of IBM Data Lakehouse (Watsonx.data)**. IBM Research Accomplishment 2023.

Achievements
- Distinguished PC award for the ICDE Research Track 2026.
- Outstanding technical achievement award from IBM Research for the delivery of query optimizer in Watsonx.data 2.0 and order of magnitude performance improvement for TPC-DS 100 TBytes on Fusion HCI.
- First time patent application invention achievement award from IBM for a patent on active learning-based ontology alignment.
- Best reviewer award for the VLDB 2023 PhD Workshop.
- Engineering Grad Fellowship from the Ira A. Fulton School of Engineering and the Polytechnic School at Arizona State University for the academic years 2018-19 and 2021-22.
- SIGMOD 2017 Student Travel Awards from the conference committee and ASU GPSA.
- Research Scholarship from the National University of Singapore.
- All India Rank of 570 in the 2007 Computer Science GATE exam.
- Undergraduate Merit Scholarship from Vellore Institute of Technology in 2007 for best academic performance.

Publications Venkata Vamsikrishna Meduri, David Kreismann, Ronald Barber, Berthold Reinwald: PTO: A Workload-driven Predictive Table Optimizer for Lakehouse Systems. ACM SIGMOD 2026 (research track)

Bastian Lipka, Venkata Vamsikrishna Meduri, Berthold Reinwald, Nasrullah Sheikh: Schema-GraphRAG: Bridging Hybrid Search and Graph Traversal for Complex Retrieval Tasks. IEEE ICDE 2026 (demo track)

Venkata Vamsikrishna Meduri, Abdul Quamar, Chuan Lei, Xiao Qin, Berthold Reinwald: ALFA: Active Learning for Graph Neural Network-based Semantic Schema Alignment. The VLDB Journal: Special issue on Machine Learning and Databases, VLDB J. 33(4): 981-1011 (2024).

Kanchan Chowdhury, Venkata Vamsikrishna Meduri, Mohamed Sarwat: A Machine Learning-Aware Data Re-partitioning Framework for Spatial Datasets. ICDE 2022 (research track)

Setu Shah, Venkata Vamsikrishna Meduri, Mohamed Sarwat: GEM: An Efficient Entity Matching Framework for Geospatial Data. ACM SIGSPATIAL 2021 (poster paper)

Venkata Vamsikrishna Meduri, Kanchan Chowdhury, Mohamed Sarwat: Evaluation of Machine Learning Algorithms in Predicting the Next SQL Query From the Future. ACM TODS 2021, Volume 46, Issue 1, Article No.: 4

Venkata Vamsikrishna Meduri, Lucian Popa, Prithviraj Sen, Mohamed Sarwat: A Comprehensive Benchmark Framework for Active Learning Methods in Entity Matching. ACM SIGMOD 2020 (research track)

Venkata Vamsikrishna Meduri, Kanchan Chowdhury, Mohamed Sarwat: Recurrent Neural Networks for Dynamic User Intent Prediction in Human-Database Interaction. EDBT 2019 (short paper)

Naser Ahmadi, Viet-Phi Huynh, Venkata Vamsikrishna Meduri, Stefano Ortona, Paolo Papotti: Mining Expressive Rules in Knowledge Graphs. ACM JDIQ Special Issue on Quality Assessment of Knowledge Graphs 2019

Mohamed Sarwat, Venkata Vamsikrishna Meduri: Human-Database Interaction - A Holistic Approach. ICDE 2018 (lightning talk abstract)

Paolo Papotti, Stefano Ortona, Venkata Vamsikrishna Meduri: RuDiK: Rule Discovery in Knowledge Bases. PVLDB 2018 (demo track)

Stefano Ortona, Venkata Vamsikrishna Meduri, Paolo Papotti: Robust Discovery of Positive and Negative Rules in Knowledge-Bases. ICDE 2018 (research track)

Rohit Singh, Venkata Vamsikrishna Meduri, Ahmed Elmagarmid, Samuel Madden, Paolo Papotti, Jorge-Arnulfo Quiane-Ruiz, Armando Solar-Lezama, Nan Tang: Synthesizing Entity Matching Rules by Examples. PVLDB 2017 (research track)

Rohit Singh, Venkata Vamsikrishna Meduri, Ahmed Elmagarmid, Samuel Madden, Paolo Papotti, Jorge-Arnulfo Quiane-Ruiz, Armando Solar-Lezama, Nan Tang: Generating Concise Entity Matching Rules. ACM SIGMOD 2017 (demo track)

Venkata Vamsikrishna Meduri, Paolo Papotti: Towards User-Aware Rule Discovery. ISIP 2016

Sushovan De, Yuheng Hu, Venkata Vamsikrishna Meduri, Yi Chen, Subbarao Kambhampati: BayesWipe: A Scalable Probabilistic Framework for Cleaning BigData. ACM JDIQ Special Issue on Web Data Quality 2016

Lydia Manikonda, Venkata Vamsikrishna Meduri & Subbarao Kambhampati: Tweeting the Mind and Instagramming the Heart: Exploring Differentiated Content Sharing on Social Media. ICWSM 2016 (short paper)

Tathagata Chakraborti, Venkata Vamsikrishna Meduri, Vivek Dondeti & Subbarao Kambhampati: A Game Theoretic Approach to Ad-hoc Coalitions in Human-Robot Societies. AAAI MIPC 2016

Venkata Vamsikrishna Meduri, Zhan Su, Kian-Lee Tan: A Write Efficient PCM-Aware Sort. DEXA (1) 2012: 86-100.

Venkata Vamsikrishna Meduri, Kian-Lee Tan: Subquery Plan Reuse based Query Optimization. COMAD 2011: 35-46.

Patents

Berthold Reinwald, David Kreismann, Ronald J. Barber, Venkata Vamsikrishna Meduri: Workload-driven Predictive Optimization of Lakehouse Table Parameters. US Patent App. 19/530,845.

Abdul Quamar, Xiao Qin, Berthold Reinwald, Venkata Vamsikrishna Meduri: Active Learning for Graph Neural Network based Semantic Schema Alignment. US Patent App. 18/191,024.
`https://patents.google.com/patent/US20240330693A1/en`

Abdul Quamar, Fatma Özcan, Chuan Lei, Vasilis Efthymiou, Venkata Vamsikrishna Meduri: Guided Exploration for Conversational Business Intelligence. US Patent App. 17/188,405.
`https://patents.google.com/patent/US20220277031A1/en`

Academic Service

PC Member

- ACM SIGMOD Research Track 2027
- PVLDB Research Track 2026, 2027
- ICDE Research Track 2026
- ICDE Industry Track 2024, 2025, 2026
- PVLDB Research Track 2025 (Rapid Response Reviewer)
- EDBT Industry Track 2024, 2025
- ACM Symposium on Cloud Computing (SoCC) 2023
- VLDB Industry Track 2023
- VLDB PhD Workshop 2023
- Tabular Data Analysis (TaDA) Workshop @ VLDB 2023, 2026

Journal Reviewer

- ACM TODS
- VLDB Journal
- Neural Networks (NEUNET)
- Journal of Web Semantics (JoWS)
- ICT Express
- Information Sciences
- International Journal of Machine Learning and Cybernetics (IJMLC)

External Reviewer

- The Web Conference (formerly known as WWW) Industry Track 2024 (served as a subreviewer)
- VLDB Research Track 2018-2020, 2023, 2024
- SIGMOD 2018
- SIGSPATIAL 2018
- HILDA @ SIGMOD 2019
- ICDE 2020