

## News Release

**Contact:** Jim Ormond  
212-626-0505  
ormond@hq.acm.org

### **ACM Prize in Computing Honors Matei Zaharia for Foundational Contributions to Data and Machine Learning Systems**

*Open-Source Systems, Including Apache Spark, Delta Lake, and MLflow Redefined Data Processing and Enabled Modern AI at Scale*

**NEW YORK – April 8, 2026** – ACM, the Association for Computing Machinery, today named Matei Zaharia as the recipient of the [ACM Prize in Computing](#) for his visionary development of distributed data systems and computing infrastructure, which has enabled large-scale machine learning, analytics, and AI at global scale.

The ACM Prize in Computing recognizes early-to-mid-career computer scientists whose work has had broad and lasting impact. The award carries a \$250,000 prize, with financial support provided by an endowment from Infosys Ltd, a global leader in next-generation digital services and consulting.

Zaharia’s work addressed a central challenge in computing: how to work with and analyze rapidly growing volumes of data efficiently, and at a scale previously accessible only to the largest technology companies. Early distributed data systems were limited in speed and poorly suited to emerging workloads such as machine learning and interactive analysis. Through a sequence of open-source systems, each targeting a distinct bottleneck, Zaharia changed what any organization could do with massive datasets.

As a doctoral student at UC Berkeley, Zaharia started Apache Spark, a new approach to distributed computing that reliably leverages memory to accelerate computations. This design made Spark dramatically faster than existing frameworks for the kinds of iterative computations essential to machine learning, while its unified architecture allowed batch processing, streaming, graph computation, and interactive queries to run within a single system. Spark quickly moved from research into widespread use and is now the de facto standard for large-scale data analytics, deployed across tens of thousands of organizations and integrated into major cloud platforms. Zaharia’s doctoral dissertation on Spark received the [ACM Doctoral Dissertation Award in 2014](#).

With the shift to the cloud, Zaharia turned to a different problem: the lack of reliability and consistency in sprawling cloud data lakes – or the massive, centralized, and often unmanaged repositories storing vast amounts of raw data. He co-developed Delta Lake to bring transactional guarantees and principled data management to cloud object stores, making data pipelines more dependable and enabling a new class of architecture – the data lakehouse – that combines the flexibility of data lakes with the reliability of traditional data warehouses. Delta Lake is now widely adopted across industries, handling exabytes of data daily.

The growing use of machine learning introduced additional complexity. Zaharia started MLflow, another open-source platform to address fragmentation in machine learning and AI workflows, where teams struggled to track experiments, reproduce results, and deploy models consistently. MLflow provided a structured framework for managing the machine learning lifecycle – from experiment tracking and model versioning to deployment across diverse tools and environments – and has become a leading platform for operationalizing AI at scale. Together, these systems reshaped how data is leveraged in practice.

By building tools that any organization could freely use and extend, Zaharia ensured that the benefits of scalable computing became accessible to researchers, nonprofits, and enterprises across every industry. As investment in artificial intelligence accelerates, the infrastructure he built remains key to how data is processed, managed, and used to train and deploy AI applications and agents.

Today, Zaharia is focusing research on AI development, specifically how to build and scale reliable agents. He is a co-author on recent open source research, including [DSPy](#) and [GEPA](#), which focus on auto-optimizing prompts and models to improve agent quality for specific tasks.

“Matei Zaharia’s work has had a lasting impact on how data is used at scale,” said ACM President Yannis Ioannidis. “By addressing key limitations in earlier systems, he developed technologies that quickly became standard tools for data analytics, machine learning, and artificial intelligence. Matei’s open-source philosophy has been essential: he made these tools accessible to all. His contributions continue to influence both research and industry, and I look forward to seeing where his current work on AI systems takes us next.”

Salil Parekh, Chief Executive Officer, Infosys, said, “Matei Zaharia’s contributions have helped define how organizations work with data and AI today. His systems are widely used across industries and have enabled teams to build, deploy and scale AI applications more effectively. Infosys is proud to support the ACM Prize in Computing since its origination in 2007.”

### **Biographical Background**

Matei Zaharia is an Associate Professor of EECS at the University of California, Berkeley, and a Cofounder and CTO of Databricks. He started the Apache Spark open-source project during his PhD at UC Berkeley in 2009, and has worked broadly on other widely used data and AI software, including Delta Lake, MLflow, Dolly and ColBERT. He currently works on a variety of research projects in cloud computing, database management, AI and information retrieval. Zaharia’s honors include the 2014 ACM Doctoral Dissertation Award, an NSF CAREER Award, the SIGOPS Mark Weiser Award, and the US Presidential Early Career Award for Scientists and Engineers (PECASE).

Zaharia will be formally presented with the ACM Prize in Computing at ACM’s annual Awards Banquet, which will be held on Saturday, June 13 at The Palace Hotel in San Francisco.

### **About the ACM Prize in Computing**

The [ACM Prize in Computing](#) recognizes an early to mid-career fundamental innovative contribution in computing that, through its depth, impact, and broad implications, exemplifies the greatest achievements in the discipline. The award carries a prize of \$250,000. Financial support is provided by an endowment from Infosys Ltd.

### **About ACM**

[ACM, the Association for Computing Machinery](#), is the world's largest educational and scientific computing society, uniting computing educators, researchers, and professionals to inspire dialogue, share resources, and address the field's challenges. ACM strengthens the computing profession's collective voice through strong leadership, promotion of the highest standards, and recognition of technical excellence. ACM supports the professional growth of its members by providing opportunities for life-long learning, career development, and professional networking.

###