



# **ACM Fellow Distinguished Lecture**

Taming Big Data: Stream Summarization and its Many Applications

**Amr El Abbadi**

University of California at Santa Barbara

**Hosts: Shantanu Sharma and Vincent Oria**

**Date:** Friday, November 21, 2025

**Coffee:** 11:15 AM – 11:30 AM

**Time:** 11:30 AM – 12:30 PM

**Location:** GITC 1400 (1st floor Lecture Hall)

**Zoom Link:** <https://njit-edu.zoom.us/j/95903616720?pwd=UrsFtzLWaZFP0ePqVmqlLxAOES8ECNn.1>

## **Abstract:**

During the past two decades we have seen an unprecedented increase in the amount of data that is generated from numerous internet-scale applications. As hundreds of millions to billions of users interact with these applications, there is a continuous flow of interactions that are collected by internet companies hosting these applications. Before this data can be subject to modeling and analysis, it is often necessary to obtain summary statistics such as the cardinality of unique visitors, frequency counts of users from different states or countries, and in general, finding the quantile information from the dataset. Efficient algorithms exist for computing the exact information over the data. Unfortunately, these algorithms require a considerable amount of time, scanning the data multiple times, or require additional storage that is linear in the size of the dataset itself. Approximation methods, with guaranteed error bounds, developed in the context of streaming data are extremely effective to extract useful and relatively accurate knowledge from big data. In this talk, we will review the recent, and not so recent, advances in big data stream summarization. The main objective of this talk is to demonstrate the strong relationship between the mathematics of big data and the management of big data. We will discuss streaming data summarization focusing on the heavy hitter's problem in diverse setting, including recent advance for environments with both insertions and deletions; privacy challenges and applications for caching in large scale elastic cloud environments and data analysis and monitoring in modern software defined networks.

## **Bio:**



Amr El Abbadi is a Professor of Computer Science. He received his B. Eng. from Alexandria University, Egypt, and his Ph.D. from Cornell University. His research interests are in the fields of fault-tolerant distributed systems and databases, focusing recently on Cloud data management, blockchain based systems and privacy concerns. Prof. El Abbadi is an ACM Fellow, AAAS Fellow, and IEEE Fellow. He was Chair of the Computer Science Department at UCSB from 2007 to 2011. He served as Associate Graduate Dean at the University of California, Santa Barbara from 2021--2023. He served as a journal editor for several database journals, including The VLDB Journal, IEEE Transactions on Computers and The Computer Journal. He has been Program Chair for multiple database and distributed systems conferences, including most recently SIGMOD 2022. He served on the executive committee of the IEEE Technical Committee on Data Engineering (TCDE) and was a board member of the VLDB Endowment from 2002 to 2008. In 2007, Prof. El Abbadi received the UCSB Senate Outstanding Mentorship Award for his excellence in mentoring graduate students. In 2013, his student, Sudipto Das received the SIGMOD Jim Gray Doctoral Dissertation Award. Prof. El Abbadi is also a co-recipient of the Test of Time Award at EDBT/ICDT 2015. In 2020, Prof El Abbadi was elected Outstanding CS faculty by the graduating seniors at UCSB. Recently, papers he co-authored received an Outstanding paper award in NSDI (Networked System Design and Implementation) 2024 and the Test of Time Award from MDM (Mobile Data Management)2024. He has published over 350 articles in databases and distributed systems and has supervised over 40 PhD students.