November 18, 2020 @ 7:00pm

Big Data Analytics with Apache AsterixDB

Featuring our speaker **Michael Carey, Ph.D.  
UCI Bren Professor of Information and Computer Sciences**

|  |  |  |  |
| --- | --- | --- | --- |
| Apache AsterixDB is a Big Data Management System (BDMS) with a feature set chosen to target use cases such as web data warehousing and social media data analysis.  This talk will provide a brief user-level overview of the system, which in current industrial terminology might be classified as a "parallel NoSQL document database system".  It will then dive into how Apache AsterixDB's SQL++ language can be used to query and analyze large volumes of semistructured (JSON) data. "NoSQL" does NOT mean NoQueries! Its notable features include:   * A NoSQL-style data model based on extending JSON with object database concepts; * A declarative query language, SQL++, that supports a broad range of queries against multiple semi-structured datasets; * A query optimizer for parallel queries and an efficient dataflow execution engine for partitioned-parallel query execution; * Partitioned and LSM-based native storage and indexing for large datasets; * Support for querying of external data (e.g., data on AWS S3) as well as natively stored data; * Rich data type support, including numeric, textual, temporal, and simple spatial data; * Basic NoSQL-like transactional capabilities.   You can click on this [link](https://acm-org.zoom.us/webinar/register/6516008917850/WN_BHuLS-saSb6xJPyMqAj8Cg) to sign up for the talk and receive a link to the zoom webinar. |  | |  | | --- | | A person holding a guitar  Description automatically generated |   Michael Carey received his B.S. and M.S. degrees from Carnegie-Mellon University and his Ph.D. from the University of California, Berkeley. He is currently a Bren Professor of Information and Computer Sciences and Distinguished Professor of Computer Science at UC Irvine, where he leads the AsterixDB project, as well as a Consulting Architect at Couchbase, Inc.  Before joining UCI in 2008, he worked at BEA Systems for seven years and led the development of their AquaLogic Data Services Platform product for virtual data integration. He also spent a dozen years at the University of Wisconsin-Madison, five years at the IBM Almaden Research Center working on object-relational databases, and a year and a half at e-commerce platform startup Propel Software during the infamous 2000-2001 Internet bubble. He is an ACM Fellow, an IEEE Fellow, a member of the National Academy of Engineering, and a recipient of the ACM SIGMOD E.F. Codd Innovations Award. His current interests center around data-intensive computing and scalable data management (a.k.a. Big Data). |