

**Chen Li**

[chenli@gmail.com](mailto:chenli@gmail.com)

<https://chenli.ics.uci.edu/>

**Title:** Data Analytics as a Service for Data Scientists

**Abstract:** Data scientists and domain experts often face challenges when dealing with large amounts of data, especially due to the scale and limited IT knowledge and infrastructure maintenance skills. In this talk, I will present several software solutions we are developing to support data analytics as a service to these users. These solutions include Apache AsterixDB as an open source parallel database, Cloudberry as a middleware system to support data visualization, and Texera as a system to enable browser-based text analytics using declarative workflows. These solutions can be integrated to support data ingestion, storage, indexing, querying, visualization, and analytics. As an example, we will report experiences of using these solutions to support management of large-scale social media data (e.g., billions of tweets in terabytes) as a service to researchers of various disciplines such as social science and public health from several schools and universities.