Challenges of Big Data in Scientific Discovery

Benjamin W. Wah, Ph.D.
Provost and Wei Lun Professor of Computer Science and Engineering
The Chinese University of Hong Kong
Shatin, Hong Kong

Abstract

Big Data is emerging as one of the hottest multi-disciplinary research fields in recent years. Big data innovations are transforming science, engineering, medicine, healthcare, finance, business, and ultimately society itself. In this presentation, we examine the key properties of big data (volume, velocity, variety, and veracity) and their relation to some applications in science and engineering. To truly handle big data, new paradigm shifts (as advocated by the late Dr. Jim Gray) will be necessary. Successful applications in big data will require in situ methods to automatically extracting new knowledge from big data, without requiring the data to be centrally collected and maintained. Traditional theory on algorithmic complexity may no longer hold, since the scale of the data may be too large to be stored or accessed. To address the potential of big data in scientific discovery, challenges on data complexity, computational complexity, and system complexity will need to be solved. We illustrate these challenges by drawing on examples in various applications in science and engineering.