



WiSE Wednesday Lecture Series Presents:

Building Scalable Data Pipeline Infrastructure: Concepts and Techniques

Yin Lu, Ph.D.

Research Associate, National Wind Institute, Texas Tech University

ABSTRACT:

Given the tremendous growth of Internet of Things (IoT), data generation capabilities in wind scientific and engineering domains have grown rapidly over the past decades. The research facilities managed by the National Wind Institute (NWI) consist of a large number of IoT sensors. For example, more than one hundred West Texas Mesonet (WTM) sites are geographically distributed across west Texas, eastern New Mexico and southwest Colorado. A 200-meter meteorological tower has 42 sensors at 10 levels above ground level. These IoT sensors provide real-time atmospheric and agricultural data including wind speed and direction, temperature, relative humidity and barometric pressure. The explosion in data volume, variety, and complexity has required scalable data pipeline infrastructure to improve the sustained performance continuously.

Distributed computing platform is considered as an effective solution to handle the enormous data generated from the IoT environment. This presentation first discusses the existing IoT data challenges. Next, the data-driven workflow and distributed system architecture are introduced for data movement, analysis, and transformation. The coordinated architecture organizes the data pipeline to achieve desired parallelism, locality, and scalability. Distributed file system, non-relational database and parallel data processing model with some technical details are presented in the last part of the talk.

BIOGRAPHY:

Dr. Yin Lu is a Research Associate in the National Wind Institute of Texas Tech University. Dr. Lu received his Ph.D. in Computer Science from Texas Tech University in 2015. His research interests include scientific data management, parallel I/O architectures, programming models, runtime system optimizations, and supporting high performance computing on large-scale systems.