

# StreamSMART: Large Scale Stream Data Processing and Its Application to Real-time Big Data

Dr. Wei Fan, Associate Director of Huawei Noah's  
Ark Lab, Hong Kong

As a typical example of New Moore's law, the amount of 3G mobile broadband (MBB) data has grown from 15 to 20 times in the past two years (> 40TB per day for typical city in China by December 2012 and it is still growing), real-time processing and mining of these data are becoming increasingly necessary. The overhead of storage and file transfer to HDFS, delay in processing, etc are making offline analysis on these datasets obsolete. Analysis of these datasets are non-trivial, examples include mobile personal recommendation, anomaly traffic detection, and network fault diagnosis. In this talk, we describe StreamSMART- Self Monitoring and Automatic Recovery Technology Streaming Platform. We discuss its architectural design that considers high throughput, low latency, fault-tolerance, usability etc, as well as its applications on MBB real-time analysis. The accuracy of statistical analysis and pattern recognition by StreamSMART is identical to that of off line analysis, while StreamSMART can process data at line rate. The design and the unique features of StreamSMART is suitable for lots of real-time processing of real-time big data.

## Short Bio:

Dr. Wei Fan is the associate director of Huawei Noah's Ark Lab. Prior to joining Huawei, he received his PhD in Computer Science from Columbia University in 2001 and had been working in IBM T.J. Watson Research since 2000. His main research interests and experiences are in various areas of data mining and database systems, such as, stream computing, high performance computing, extremely skewed distribution, cost-sensitive learning, risk analysis, ensemble methods, easy-to-use nonparametric methods, graph mining, predictive feature discovery, feature selection, sample selection bias, transfer learning, time series analysis, bioinformatics, social network analysis, novel applications and commercial data mining systems. His co-authored paper received ICDM'2006 Best Application Paper Award, he lead the team that used Random Decision Tree to win 2008 ICDM Data Mining Cup Championship. He is the associate editor of ACM Transaction on Knowledge Discovery and Data Mining (TKDD).

August 19, 2013

3:30pm-4:30pm

Davis Hall 113A