# Big Data Integration

Divesh Srivastava
AT&T Labs, Research

**Abstract**

The Big Data era is upon us: data is being generated, collected and analyzed at an unprecedented scale, and data-driven decision making is sweeping through all aspects of society. Since the value of data explodes when it can be linked and fused with other data, addressing the big data integration (BDI) challenge is critical to realizing the promise of Big Data. BDI differs from traditional data integration in many dimensions: (i) the number of data sources, even for a single domain, has grown to be in the tens of thousands, (ii) many of the data sources are very dynamic, as a huge amount of newly collected data are continuously made available, (iii) the data sources are extremely heterogeneous in their structure, with considerable variety even for substantially similar entities, and (iv) the data sources are of widely differing qualities, with significant differences in the coverage, accuracy and timeliness of data provided. This talk explores the progress that has been made by the data integration community in addressing these novel challenges faced by big data integration, and identifies a range of open problems for the community.

**Biography**

Divesh Srivastava is the head of Database Research at AT&T Labs-Research. He received his Ph.D. from the University of Wisconsin, Madison, and his B.Tech. from the Indian Institute of Technology, Bombay. He is an ACM fellow, on the board of trustees of the VLDB Endowment and an associate editor of the ACM Transactions on Database Systems. His research interests and publications span a variety of topics in data management.

The 19th International Conference on Management of Data (COMAD),
19th-21st Dec, 2013 at Ahmedabad, India.