

## Industry Practice Expo Invited Talk

# Hadoop: A View from the Trenches

Milind Bhandarkar

Pivotal

Milind.Bhandarkar@emc.com

### ABSTRACT

From its beginnings as a framework for building web crawlers for small-scale search engines to being one of the most promising technologies for building datacenter-scale distributed computing and storage platforms, Apache Hadoop has come far in the last seven years. In this talk I will reminisce about the early days of Hadoop, and will give an overview of the current state of the Hadoop ecosystem, and some real-world use cases of this open source platform. I will conclude with some crystal gazing in the future of Hadoop and associated technologies.

### Categories and Subject Descriptors

H.2.8 [Database Management]: Data Mining.

### General Terms

Performance, Reliability

### Keywords

Hadoop; Open Source

### Bio

Milind Bhandarkar was the founding member of the team at Yahoo! that took Apache Hadoop from 20-node prototype to datacenter-scale production system, and has been contributing and working with Hadoop since version 0.1.0. He started the Yahoo! Grid solutions team focused on training, consulting, and supporting hundreds of new migrants to Hadoop. Parallel programming languages and paradigms has been his area of focus for over 20 years, and his area of specialization for PhD (Computer Science) from University of Illinois at Urbana-Champaign. Previously, he has worked at the Center for Development of Advanced Computing (C-DAC), National Center for Supercomputing Applications (NCSA), Center for Simulation of Advanced Rockets, Siebel Systems, Pathscale Inc. (acquired by QLogic), Yahoo! and LinkedIn. Currently, he is the Chief Scientist, Machine Learning Platforms at Pivotal.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).

*KDD'13*, August 11–14, 2013, Chicago, Illinois, USA.

ACM 978-1-4503-2174-7/13/08.