

Trends and Challenges in Big Data Processing

Ion Stoica

University of California, Berkeley

istoica@cs.berkeley.edu

ABSTRACT

Almost six years ago we started the Spark project at UC Berkeley. Spark is a cluster computing engine that is optimized for in-memory processing, and unifies support for a variety of workloads, including batch, interactive querying, streaming, and iterative computations. Spark is now the most active big data project in the open source community, and is already being used by over one thousand organizations.

One of the reasons behind Spark's success has been our early bet on the continuous increase in the memory capacity and the feasibility to fit many realistic workloads in the aggregate memory of typical production clusters. Today, we are witnessing new trends, such as Moore's law slowing down, and the emergence of a variety of computation and storage technologies, such as GPUs, FPGAs, and 3D Xpoint. In this talk, I'll discuss some of the lessons we learned in developing Spark as a unified computation platform, and the implications of today's hardware and software trends on the development of the next generation of big data processing systems.

BIOGRAPHY

Ion Stoica is a Professor in the EECS Department at University of California at Berkeley. He does research on cloud computing and networked computer systems. Past work includes the Dynamic Packet State (DPS), Chord DHT, Internet Indirection Infrastructure (i3), declarative networks, replay-debugging, and multi-layer tracing in distributed systems.

He is an ACM Fellow and has received numerous awards, including the SIGCOMM Test of Time Award (2011), and the ACM doctoral dissertation award (2001). In 2006, he co-founded Conviva, a startup to commercialize technologies for large scale video distribution, and in 2013, he co-founded Databricks a startup to commercialize Apache Spark.

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>. For any use beyond those covered by this license, obtain permission by emailing info@vlDB.org.

Proceedings of the VLDB Endowment, Vol. 9, No. 13
Copyright 2016 VLDB Endowment 2150-8097/16/09.