Title: Efficient Data Formats for Large Scale Data Analysis

Abstract:
We will review the state-of-the-art open source storage formats and APIs for efficient analysis of the heterogeneous data sets at the internet scale. We will review the implementation algorithms, and the architecture the APIs, their suitability for research and development applications, and some industrial case studies.

Bio: Dr. Edmon Begoli is a chief data officer with Joint Institute for Computational Sciences (JICS). He is the Chief Technology Officer of PYA Analytics and has primary responsibility for technology vision, strategy and research, and development of analytic products and solutions.

Edmon has been a technical architect for more than 17 years, most recently serving as a chief architect at Oak Ridge National Laboratory for the nationally recognized Knowledge Discovery Infrastructure (KDI) for the Centers for Medicare & Medicaid Services (CMS). Prior to his position at ORNL, Edmon served in various technology leadership roles at the Y-12 National Security Complex, Laboratory Corporation of America and Lucent Technologies. His work on large scale analytic architectures has been published in numerous international professional and research publications. Edmon holds a bachelor’s, master’s (University of Colorado-Boulder) and PhD (University of Tennessee-Knoxville) degrees in computer science.