EXTREME DATA

8  Guest Editors’ Introduction
   Manish Parashar and George K. Thiruvathukal
   Extreme Data

11  Big Data Applications Using Workflows for Data Parallel Computing
    Jianwu Wang, Daniel Crawl, Ilkay Altintas, and Weizhong Li
    In the Big Data era, workflow systems must embrace data parallel computing
techniques for efficient data analysis and analytics. Here, an easy-to-use, scalable
approach is presented to build and execute Big Data applications using actor-oriented
modeling in data parallel computing. Two bioinformatics use cases for next-
generation sequencing data analysis demonstrate the approach’s feasibility.

22  Ten Years of SkyServer I: Tracking Web and SQL e-Science Usage
    M. Jordan Raddick, Ani R. Thakar, Alexander S. Szalay, and Rafael D.C. Santos
    SkyServer is the primary catalog data portal of the Sloan Digital Sky Survey that
makes multiple terabytes of astronomy data available to the world. Here, the process
is described of collecting and analyzing the complete record of more than 10 years of
Web hits and SQL queries to SkyServer.

32  Ten Years of SkyServer II: How Astronomers and the Public Have
    Embraced e-Science
    M. Jordan Raddick, Ani R. Thakar, Alexander S. Szalay, and Rafael D.C. Santos
    A comprehensive analysis of 10 years of Web and SQL traffic on SkyServer—the online
portal to the multiterabyte Sloan Digital Sky Survey archive—shows the impressive reach
of the SDSS to the research community and the public, and provides insight into how
methods of e-science are being taken up by the scientific community.

41  IK-SVD: Dictionary Learning for Spatial Big Data via Incremental
    Atom Update
    Lizhe Wang, Ke Lu, Peng Liu, Rajiv Ranjan, and Lajiao Chen
    Big Data, a large and complex collection of datasets characterized by four V’s
(volume, variety, veracity, and velocity), is difficult to deal with using traditional
data processing algorithms and models. A proposed dictionary learning algorithm,
which extends the classical method that uses the K-means and Singular Value
Decomposition (K-SVD) algorithm by incrementally updating atoms, will ably
represent the spatiotemporal remote sensing of Big Data and do so both efficiently
and sparsely.

53  Evaluating Geospatial Geometry and Proximity Queries Using Distributed
    Hash Tables
    Matthew Malensek, Sangmi Pallickara, and Shrideep Pallickara
    This article explores the challenges associated with supporting geospatial retrievals
constrained by arbitrary geometric bounds, geographic proximity, and relevance
rankings. The proposed solution involves the use of the geovailability grid, a
lightweight, distributed spatial indexing structure. The index can also be used in
nondistributed settings, and performs competitively with other spatial indexing
technologies.
ALSO IN THIS ISSUE

62 Federated Computing for the Masses—Aggregating Resources to Tackle Large-Scale Engineering Problems
Javier Diaz-Montes, Yu Xie, Ivan Rodero, Jaroslav Zola, Baskar Ganapathysubramanian, and Manish Parashar

An exploration of the use of aggregated high-performance computing resources to solve large-scale engineering problems shows that it’s possible to build a computational federation that’s easy for users to implement, and is elastic, resilient, and scalable. The fusion of federated computing and real-life engineering problems is brought to the average user by providing relevant middleware.

COLUMNS

4 From the Editors
Keeping Pace with Extreme Data
Judith Bayard Cushing

96 The Last Word
Control and Creativity
Charles Day

DEPARTMENTS

6 Books
Applied Mathematics: Methods and Matlab
Brandon G. Cook

73 Your Homework Assignment
NASA Computational Case Study: Modeling Planetary Magnetic and Gravitational Fields
David G. Simpson and Adolfo F. Viñas

80 Scientific Programming
Practically Accurate Floating-Point Math
Neil Toronto and Jay McCarthy

RESOURCES

Inside Front Cover
IEEE Computer Society Information
21 AIP Membership Information