Hadoop. TS: Large-Scale Time-Series Processing

Abstract

The paper describes a computational framework for time-series analysis. It allows rapid prototyping of new algorithms, since all components are re-usable. Generic data structures represent different types of time series, e.g., event and interevent time series, and define reliable interfaces to existing big data. Standalone applications, highly scalable MapReduce programs, and User Defined Functions for Hadoop-based analysis frameworks are the major modes of operation. Efficient implementations of univariate and bivariate analysis algorithms are provided for, e.g., long-term correlation, crosscorrelation and event synchronization analysis on large data sets.

References

- M. Small, "Applied nonlinear time series analysis: applications in physics,"
- J. W. Kantelhardt, "Fractal and multifractal time series," in
2012.

Index Terms

Computer Science

Algorithms

Keywords

Time Series Analysis, Detrended Fluctuation Analysis, Return Interval Statistics
Cross Correlation

Event Synchronization

Hadoop

MapReduce