

A Scalable Media Job Framework for an Open Source Search Engine

Authors : Pooja Mishra, Chris Pollett

Abstract : This paper explores efficient ways to implement various media-updating features like news aggregation, video conversion, and bulk email handling. All of these jobs share the property that they are periodic in nature, and they all benefit from being handled in a distributed fashion. The data for these jobs also often comes from a social or collaborative source. We isolate the class of periodic, one round map reduce jobs as a useful setting to describe and handle media updating tasks. As such tasks are simpler than general map reduce jobs, programming them in a general map reduce platform could easily become tedious. This paper presents a MediaUpdater module of the Yioop Open Source Search Engine Web Portal designed to handle such jobs via an extension of a PHP class. We describe how to implement various media-updating tasks in our system as well as experiments carried out using these implementations on an Amazon Web Services cluster.

Keywords : distributed jobs framework, news aggregation, video conversion, email

Conference Title : ICCTS 2016 : International Conference on Collaboration Technologies and Systems

Conference Location : New York, United States

Conference Dates : June 06-07, 2016