



TEXAS TECH UNIVERSITY

Edward E. Whitacre Jr.
College of Engineering

Computer Science

Data-Intensive Scalable Computing Laboratory (DISCL)

Technical Report

Department of Computer Science

Texas Tech University

Domino: A Trigger-based Incremental Programming Model with Non-blocking
Synchronization for Cloud Computing

Dong Dai, Xuehai Zhou, Dries Kimpe, Rob Ross, Yong Chen

{dong.dai, yong.chen}@ttu.edu xhzhou@ustc.edu
{dkimpe, rross}@mcs.anl.gov

2013-10-01

Technical Report № TTU/DISCL-2013-03

<http://discl.cs.ttu.edu>

LIMITED DISTRIBUTION NOTICE: This report has been submitted for publication outside of TTU-DISCL and will probably be copyrighted if accepted for publication. It has been issued as a Technical Report for early dissemination of its contents. In view of the transfer of copyright to the outside publisher, its distribution outside of TTU-DISCL prior to publication should be limited to peer communications and specific requests. After outside publication, requests should be filled only by reprints or legally obtained copies of the article (e.g. payment of royalties).

Domino: A Trigger-based Incremental Programming Model with Non-blocking Synchronization for Cloud Computing

Dong Dai, Xuehai Zhou, Dries Kimpe, Rob Ross, Yong Chen

{dong.dai, yong.chen}@ttu.edu xhzhou@ustc.edu

{dkimpe, rross}@mcs.anl.gov

Texas Tech University

University of Science and Technology of China

Argonne National Lab

Abstract

Many cloud applications need to process large scale online data sets that evolve over time as new entries are added and existing entries are modified. Several programming models and associated runtime systems, such as Percolator or Oolong, are designed for such incremental data processing and can achieve efficient updates by an event-driven abstraction supported by observers and triggers. However, these programming models and runtime systems either lack the desired synchronization among parallel observers/triggers or simply leave the heavy burden of managing synchronizations to applications developers. Such a limitation significantly restricts the usability and ubiquity of trigger-based incremental programming models. In this study, we propose a trigger-based programming model and associated runtime system, namely Domino, based on a popular Cloud storage system (HBase) with a self-managed eventual synchronization mechanism and runtime optimizations to coordinate parallel triggers efficiently. Use cases and extensive evaluation results confirm that the self-managed eventual synchronization and the new Domino programming model are highly efficient for incremental processing applications and much easier to use as well.

Keywords: *Cloud Computing, Programing model, Incremental Processing, Synchronization*