International Journal of Distributed Artificial Intelligence
Volume 14 • Issue 1

Distributed Business Process Discovery in Cloud Clusters

Ishak H. A Meddah, University of Saida Dr. Moulay Tahar, Algeria
https://orcid.org/0000-0003-4164-7487

Fatiha Guerroudji, USTO MB University, Algeria

Nour Elhouda Remil, University of Saida Dr. Moulay Tahar, Algeria

ABSTRACT

The processing of big data across different axes is becoming more difficult, and the introduction of the Hadoop MapReduce framework seems to be a solution to this problem. With this framework, large amounts of data can be analyzed and processed. It does this by distributing computing tasks between a group of virtual servers operating in the cloud or a large group of devices. The mining process forms an important bridge between data mining and business process analysis. Its techniques make it possible to extract information from event reports. The extraction process generally consists of two phases: identification or discovery and innovation or education. The first task is to extract small patterns from the log effects. These templates represent the implementation of the tracking from a business process report file. In this step, the authors use the available technologies. Patterns are represented by finite state automation or regular expressions. And the final model is a combination of two different styles.

KEYWORDS

Business Process, Execution Traces, Log File, Process Mining

INTRODUCTION

Process Mining (W. van der Aalst, 2016) is a new scientific discipline that enables the extraction of knowledge from data occurring in the modern data industry, for example, records historical completion data. Mining techniques are typically used to discover, monitor, and refine processes by applying multiple processes to the production of event logs by completion (W. van der Aalst et al, 2012). Process exploration, relationship testing and process improvement are the three main functions of the mining process, which are widely used in industrial processes in many applications such as finance, transportation and healthcare. Several authors in the literature on the application of mining techniques in production processes (Demtroeder et al., 2019) (Lechner, 2020). compare other application areas such as sales, purchasing, marketing and insurance. In addition, existing work focuses on the application of the data from the solution (Reinkemeyer, 2020).

DOI: 10.4018/IJDAI.301213 *Corresponding Author

Copyright © 2022, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

Many strategies have been proposed to remove these images from performance. However, most existing processes only delete simple templates or templates that are limited to only a specific set of manually selected events. Recent studies have shown that patterns can be identified in a language (Ammons et al, 2002). In this way, patterns can be seen as normal or in restricted mode, and mining patterns can be seen as a language learning problem.

The current method is of equal importance. Each uses dynamic effects or metafiles as input to the operating system and creates one or more internal languages that represent the model or function. However, personal solutions differ in the main way.

In this article, all the new methods of removing patterns that address the various limitations of the current process are presented. the understanding is twofold. First of all, it can be seen that the smaller models can be likened to larger models. Second, micro-model configurations can be integrated. Then using that knowledge to divide the job into two parts. First, techniques to isolate two of the smaller models are used. They are integrated into a finite state automaton algorithm and some special rules used by M. Gabel and Z.Su (Gabel et al, 2008) extractable by token algorithm. (Gabel & Su,., 2008).

Finally, the Hadoop MapReduce framework is used to extract and create models. These models are presented as finite state automata or continuous expressions. This chapter uses the same token extraction algorithm, however, subtracts the small sample in parallel with the map pitch and counts the small sample by the large sample in parallel with the reduced pitch. the mission is to improve the isolation models and procedures that can be used in the onboarding process.

This method is used in Java programming language with two log files for the two programs. The size of the first data wheel is 20 GB and the size of the second data is 30 GB, which is made by the manufacturer of the wheel. The 3-group method in the cloud are tested and provided with the first group of 5 machines, a second group of 10 machines, and a third group of 20 more machines. The effectiveness of the service is call, response, message ... another one.

RESEARCH COMPONENTS

Due to new computational challenges, High Performance Computing (HPC) is becoming more and more important in the analysis of business process. HPC typically involves the distribution of work among a group of devices that have access to a shared file system hosted on a storage area network. This work takes place in parallel via software Application Programming Interface (APIs) such as the Message Passing Interface (MPI) and, more recently, the Hadoop MapReduce API. The pattern of architecture and other computing services is explored, cloud computing. In short, cloud computing is equivalent to the HPC web interface for scaling up quickly and scaling for on-demand use. The server side runs in data centers that operate in clusters, and remote clients load large potential datasets for analysis in Hadoop or other parallel environments running in the data center.

MapReduce

Hadoop is a set of open source software tools that make it easy to use multiple computer networks to solve problems involving large amounts of data and computing. It provides a software framework for storing and processing big data using the MapReduce programming model.

The first step of Hadoop is the Hadoop Distributed File System (HDFS), which is used to store data, and the MapReduce programming model, which is used for distributed computing. The second stage of Hadoop will be developed in more detail by breaking down the previous MapReduce resource management function into yarn and introducing other applications such as Hive that are used to run queries. In the third step, more details are added and there are hundreds of applications in the Hadoop framework that are used for machine learning, flow data analysis, cloud environment, etc (White, 2012).

Hadoop (White, 2012) (Hadoop,2021) is a software framework that can be installed on a standard Linux stack to enable analysis of large distributed data. Other than possible changes to meet the minimum recommended Random Access Memory (RAM), disk space, etc. for each node, no

16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/distributed-business-process-discovery-in-cloud-clusters/301213

Related Content

Multi-Agent Negotiation Paradigm for Agent Selection in B2C E-Commerce
Bireshwar Dass Mazumdar, Swati Basakand Neelam Modanwal (2012). *Theoretical and Practical Frameworks for Agent-Based Systems (pp. 164-186).*
www.irma-international.org/chapter/multi-agent-negotiation-paradigm-agent/66030

Tropos: A Requirements-Driven Methodology for Agent-Oriented Software
Paolo Giorgini, Manuel Kolp, John Mylopoulosand Jaelson Castro (2005). *Agent-Oriented Methodologies (pp. 20-45).*
www.irma-international.org/chapter/tropos-requirements-driven-methodology-agent/5054

Inductive Logic Programming and Embodied Agents: Possibilities and Limitations
Andrea Kulakov, Joona Laukkanen, Blerim Mustafaand Georgi Stojanov (2009). *International Journal of Agent Technologies and Systems (pp. 34-49).*
www.irma-international.org/article/inductive-logic-programming-embodied-agents/1390

Challenges of the "Global Understanding Environment" Based on Agent Mobility
Vagan Terziyan (2007). *Application of Agents and Intelligent Information Technologies (pp. 121-152).*
www.irma-international.org/chapter/challenges-global-understanding-environment-based/5112

Communications for Agent-Based Human Team Support
Gita Sukthankar, Katia Sycara, Joseph A. Giampapaand Christopher Burnett (2009). *Handbook of Research on Multi-Agent Systems: Semantics and Dynamics of Organizational Models (pp. 285-313).*
www.irma-international.org/chapter/communications-agent-based-human-team/21104