Applied Soft Computing 38 (2016) 1049–1059

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Scheduling big data applications within advance reservation framework in optical grids

Mohamed Abouelela*, Mohamed El-Darieby

University of Regina, Regina, SK, Canada

ARTICLE INFO

Article history:
Received 30 December 2014
Received in revised form 5 July 2015
Accepted 18 August 2015
Available online 28 August 2015

Keywords:
Advance reservation
Big data
Co-scheduling
Hierarchical architecture
Multi-domain networks
Optical grids

ABSTRACT

Providing QoS for big data applications requires a way to reserve computing and networking resources in advance. Within advance reservation framework, a multi-domain scheduling process is carried out in a top down hierarchical way across multiple hierarchical levels. This ensures that each domain executes intra-domain scheduling algorithm to co-schedule its own computing and networking resources while coordinating the scheduling at the inter-domain level. Within this process, we introduce two algorithms: iterative scheduling algorithm and *K-shortest paths* algorithm. We conducted a comprehensive performance evaluation study considering several metrics that reflect both grid system and grid user goals. The results demonstrated the advantages of the proposed scheduling process. Moreover, the results highlight the importance of using the iterative scheduling and *K-shortest paths* algorithms especially for data intensive applications.

1. Introduction

Data is a major asset for organizations in the 21st century. In fact, nowadays data is actually considered an intellectual capital in financial books of companies. Newer applications, such as smart cities, e-science, are expected to have a large impact on the efficiency, safety, and sustainability of our lives and our enterprises. Traditionally, most enterprises have used data stored in a self-contained repository which results in tight-coupling of data and corresponding software. With the proliferation of the service-oriented architecture, the Data as a Service technology became to provide easier interfacing and access to data. In fact, DaaS enables data to be shared among data providers regardless of where that data came from. This allows software programmers to mashup data from different pools regardless of location.

Big data applications require collecting and transferring massive volumes of data, reaching petabytes range, from geographically distributed instruments to be processed at distributed sites [1]. Many applications, such as particle physics experiments, image processing, and linear algebra produces large amounts of data that cannot be stored in one location nor processed in reasonable time.

Collaboration among different institutions and organizations by sharing data and resources is essential to fulfil the requirements of such Big Data applications. Institutions located at geographically distributed locations may be interested in the data collected from distributed instruments such as sensors, cameras and satellites. The institutions could agree on building e-science grid environment to share data and resources. The data is transferred to storage and processing locations at different sites using high speed optical links [2].

Distributed grid systems are developing from *best effort* service towards *differentiated service* with guaranteed Quality of Service (QoS) [3]. The *best effort* behaviour does not guarantee where or when the data will be stored or processed. This behaviour could not be accepted in more complex applications such as deadline driven and real time applications [4]. Providing QoS requires resource advance reservation [3]. Advance reservation (AR) guarantees that a certain resource is available when application requires it, which ensures satisfying its QoS requirements. Advance reservation frameworks have been studied in different projects [1,5–7].

Considering the multi-domain multi-level nature of grid systems adds extra complexity to resource scheduling within AR framework. Distributed grid systems are multi-domain environments having, usually, two levels of hierarchy. In the lower level, different sites containing computing and storage resources are grouped into domains. Each domain is managed by a local domain grid manager. A global resource manager (RM), usually called meta-scheduler, is usually used to manage the negotiation and sharing among different domains.

* Corresponding author.
E-mail address: Mohamed.Abouelela@uregina.ca (M. Abouelela).

http://dx.doi.org/10.1016/j.asoc.2015.08.032
1568-4946/

In [8,9], we introduced multi-domain hierarchical scheduling process within advance reservation framework. The scheduling process is carried out in a top down hierarchical way across multiple hierarchical levels. This ensures that each domain executes intra-domain scheduling algorithms to schedule its own resources, while coordinating the scheduling at the inter-domain level.

In this paper, we extend the process by introducing two new algorithms: iterative scheduling algorithm and *K-shortest paths* algorithm. The iterative scheduling algorithm is introduced to handle more complex data intensive divisible load applications. Moreover, *K-shortest paths* algorithm is introduced for path computing. The *K-shortest paths* algorithm selects the path with minimum waiting time among k shortest paths in order to minimize data transfer delays.

The rest of the paper is organized as follows: related work is summarized in Section 2. The system components and the proposed iterative scheduling algorithm is described in Section 3.1. Experiments setup is explained in Section 4, while results and discussions are provided in Section 5. Finally, conclusions are offered at the end of the paper.

2. Related work

In this section, different projects supporting advance reservation in multi-domain environment are reviewed, followed by a summary of scheduling techniques within those frameworks. A review of scheduling divisible load applications in AR framework is provided at the end of this section.

2.1. Advance reservation in multi-domain optical grid environment

Advance reservation frameworks in optical grids were presented in different projects. OSCARS and DRAC [1] projects considered single domain environments, while EnLIGHTened [5], G-lambda [6] and Phosphorus [7] projects considered multiple domain environments. Those three projects provide similar architectures and general concepts to reserve, allocate, and monitor computing and networking resources. In this section, we will briefly discuss these projects.

The objective of the United States EnLIGHTened Computing project [5] is to design framework that allocates different types of resources such as computing, storage and optical network paths to grid applications. It supports both in advance and on demand reservation. The EnLIGHTened architecture introduced the EnLIGHTened Resource Broker (ERB). The ERB consists of a set of Highly-Available Resource Co-allocator (HARC) acceptors that manage the co-allocation process and a set of HARC resource managers that manage the reservation of different types of resources. ERB gets the resources availability information from local domain managers.

The architecture of the Japans G-lambda project [6] is similar to the EnLIGHTened project. It consists of Grid Resource Scheduler (GRM) (similar to the ERB in the EnLIGHTened project) and a set of resource managers (RMs) for computing (CRM), network (NRM), belonging to different domains. The GRM receives the resource reservation requests and co-allocates the suitable resources after the co-ordination with different RMs. The G-lambda project focuses on defining standard web services interface between the GRS and RMs.

The objective of Phosphorus project [7] is to provide on-demand and in-advance end to end provisioning of computing and networking resources in multi-domain and multi-technology environments. The architecture consists of the inter-domain broker (IDB) (similar to ERB and GRM in EnLIGHTened and G-lambda projects respectively) and Network Resource Provisioning System (NRPS) modules, similar to CRM and NRM in G-lambda project. The organization of IDBs can be done in centralized (one central IDB to allocate the inter-domain path, managing authentication, and so on) and distributed manner (multiple IDBs).

2.2. Resource scheduling in AR projects

Different resource scheduling techniques were proposed within the advance reservation framework. Harmony [10] is the network resource brokering system in Phosphorus project. The workflow when a grid task is received is as follows. After the authentication, the availability of the requested resources is verified. Then, the end to end path is allocated in two phases. In the first phase, the inter-domain path is selected by the IDB module. In the second phase, the NRPS module in each independent domain calculates the intra-domain path. The proposed Harmony system for multi-domain network resource reservation is promising. It shows how different domains can interact to provide end to end connectivity.

Lightpath reservation process in EnLIGHTened project was presented in [11]. The ERB coordinates with local domain managers to handle inter-domain requests. It conducts inter-domain coarse path computation and resource allocation functions. The inter-domain resource reservation process starts by sending the type and time range for each requested resource to the ERB. The ERB gets the availability information for different resources from local domain resource managers. According to the collected information, the ERB selects the optimal resource set and makes the reservations. Up on receiving the confirmation from the local mangers, the selected resource set returned to the requesting application.

2.3. Scheduling divisible load applications within AR framework

Generally speaking, resource scheduling for data intensive applications within advance reservation framework should provide two functionalities: *co-allocation* and *network aware* functionalities [12]. *Co-allocation* of computing and networking resources ensures the availability of the needed resources for data transfer and execute tasks. Different contributions [13–16] were presented providing co-allocation for different types of grid application such as workflow and pipelines applications.

Network awareness ensures the precision of the scheduling decisions by considering data transfer delays while deciding on resource scheduling. Traditional schedulers used to schedule computing resources without considering data transfer delays. This is acceptable in applications where data transfer delays are negligible compared to job execution time. However, as the applications are becoming more data intensive, the schedulers should consider transfer delay while deciding on scheduling computing and networking resources [17].

Designing network aware schedulers for divisible data load applications has been discussed in different contributions. In [18], the authors use the divisible load theory to solve a combined lambda grid dimensioning and scheduling problem. In [2], a GA based approach that co-schedules computational and networking resources, while considering network resources availability and connectivity was introduced. The main drawback of this approach is the long GA execution time. In [19], a greedy algorithm based on divisible load theory is introduced for single domain environment. The proposed algorithm is simple, light-weight, fast scheduler. In this paper, we extended this work by developing an iterative and incremental scheduling algorithm based on this greedy algorithm.