DISRAY: A distributed ray tracing by map-reduce

Afsaneh Mohammadzaheri, Hossein Sadeghi*, Sayyed Keivan Hosseini, Mahdi Navazandeh
Earthquake Research Center, Ferdowsi University of Mashhad, Mashhad 9177948974, Iran

Abstract
Seismic ray tracing is one of the most important tools in seismology which is required in many applications such as body wave tomography and exploration seismology. Due to the large amount of data introduced in global tomography, most studies are still using one-dimensional (1D) ray tracing methods. Furthermore, since three-dimensional (3D) global studies generally use initial models obtained by 1D ray tracing algorithms, such algorithms are essentially desired. Regarding new advances in High Performance Computing (HPC) as well as the importance of 1D models in global tomography, this study tries to improve the 1D ray tracing algorithm through an innovative programming paradigm, Map-Reduce. Based on this, we present a novel algorithm, DISRAY, for the distributed implementation of 1D ray tracing method, with the use of Map-Reduce idea. “Map” and “Reduce” are two built-in functions in Python (Python 1.4, 1996), with the following syntax:

map(function, sequence): calls function (item) for each of the sequence's items and returns a list of the return values.

reduce(func, sequence): returns a single value constructed by calling the binary function func on the first two items of the sequence, then on the result and the next item, and so on.

In several branches of seismology and seismic exploration, there is a need for fast and precise ray tracing to find the paths that seismic waves travel from the source to the receiver. Seismic global tomography is one of these main fields whose objective is to construct a global seismic velocity structure of the Earth (e.g., Dueker and Sheehan, 1998; Widiyantoro et al., 2000; Zhao, 2001, 2004). Since high amounts of travel time data are usually used in the global tomographic studies, they are still employing one-dimensional (1D) ray tracing methods, in which velocity varies only with depth (Zhao, 2004). Moreover, high expense and complexity of 3D ray tracing have made the implementation of 3D global tomography so difficult that most studies have generally used 1D ray tracing algorithms (Widiyantoro et al., 2000). Computing capabilities have reached to such a state that it is now essential for seismologists to develop more accurate and well-performed algorithms. Some researchers have addressed the issue; e.g., Bohlen (2002) presented a parallel finite difference algorithm for modeling 3D viscoelastic waves by MPI. Grunberg et al. (2004) designed an algorithm in which a set of seismic rays are traced in an adaptive mesh model. They tested their algorithm in three different parallel architectures. Komatitsch et al. (2010) developed a high-order finite element technique on GPUs for modeling the seismic waves in an elastic medium. Here, we present DISRAY, a distributed algorithm for the shooting ray tracing method, with the use of Map-Reduce idea. “Map” and “Reduce” are two built-in functions in Python (Python 1.4, 1996), with the following syntax:

map(function, sequence): calls function (item) for each of the sequence's items and returns a list of the return values.

reduce(func, sequence): returns a single value constructed by calling the binary function func on the first two items of the sequence, then on the result and the next item, and so on.

Later, Google researchers Dean and Ghemawat (2004) introduced a simple programming model based on the Map-Reduce idea. They employed their MapReduce to implement distributed computing on large data sets on clusters. The scale of data processed by Google's MapReduce is massively significant and Google currently uses MapReduce to process almost 1 exabyte of data every month (Menon et al., 2011). The main power of MapReduce is due to its significant simplicity (Lin and Dyer, 2010; Dean and Ghemawat, 2004). Even programmers without any knowledge of parallel and distributed computing, can easily utilize MapReduce. One of the main properties of MapReduce is that it tolerates failures in individual computers (Dean and Ghemawat, 2004). In this paradigm, distributed computing is performed only by specifying two ordinary functions, Map and Reduce. Hadoop (http://hadoop.apache.org/mapreduce/) is also

1. Introduction

In several branches of seismology and seismic exploration, there is a need for fast and precise ray tracing to find the paths that seismic waves travel from the source to the receiver. Seismic global tomography is one of these main fields whose objective is to construct a global seismic velocity structure of the Earth (e.g., Dueker and Sheehan, 1998; Widiyantoro et al., 2000; Zhao, 2001, 2004). Since high amounts of travel time data are usually used in the global tomographic studies, they are still employing one-dimensional (1D) ray tracing methods, in which velocity varies only with depth (Zhao, 2004). Moreover, high expense and complexity of 3D ray tracing have made the implementation of 3D global tomography so difficult that most studies have generally used 1D ray tracing algorithms (Widiyantoro et al., 2000). Computing capabilities have reached to such a state that it is now essential for seismologists to develop more accurate and well-performed algorithms. Some researchers have addressed the issue; e.g., Bohlen (2002) presented a parallel finite difference algorithm for modeling 3D viscoelastic waves by MPI. Grunberg et al. (2004) designed an algorithm in which a set of seismic rays are traced in an adaptive mesh model. They tested their algorithm in three different parallel architectures. Komatitsch et al. (2010) developed a high-order finite element technique on GPUs for modeling the seismic waves in an elastic medium. Here, we present DISRAY, a distributed algorithm for the shooting ray tracing method, with the use of Map-Reduce idea. “Map” and “Reduce” are two built-in functions in Python (Python 1.4, 1996), with the following syntax:

map(function, sequence): calls function (item) for each of the sequence's items and returns a list of the return values.

reduce(func, sequence): returns a single value constructed by calling the binary function func on the first two items of the sequence, then on the result and the next item, and so on.

Later, Google researchers Dean and Ghemawat (2004) introduced a simple programming model based on the Map-Reduce idea. They employed their MapReduce to implement distributed computing on large data sets on clusters. The scale of data processed by Google's MapReduce is massively significant and Google currently uses MapReduce to process almost 1 exabyte of data every month (Menon et al., 2011). The main power of MapReduce is due to its significant simplicity (Lin and Dyer, 2010; Dean and Ghemawat, 2004). Even programmers without any knowledge of parallel and distributed computing, can easily utilize MapReduce. One of the main properties of MapReduce is that it tolerates failures in individual computers (Dean and Ghemawat, 2004). In this paradigm, distributed computing is performed only by specifying two ordinary functions, Map and Reduce. Hadoop (http://hadoop.apache.org/mapreduce/) is also

*Correspondence to: Earthquake Research Center, Ferdowsi University of Mashhad, Azadi square, Mashhad 9177948974, Iran. Tel.: +98 511 879 1195; fax: +98 511 879 1191.
E-mail addresses: sadeghi@um.ac.ir, sadeghimail@yahoo.com (H. Sadeghi).

0098-3004/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cageo.2012.10.009