

Chapter 2

Introducing Data Structures for Big Data

Ranjit Biswas

Jamia Hamdard University, India

ABSTRACT

The homogeneous data structure ‘train’ and the heterogeneous data structure ‘atrain’ are the fundamental, very powerful dynamic and flexible data structures, being the first data structures introduced exclusively for big data. Thus ‘Data Structures for Big Data’ is to be regarded as a new subject in Big Data Science, not just as a new topic, considering the explosive momentum of the big data. Based upon the notion of the big data structures train and atrain, the author introduces the useful data structures for the programmers working with big data which are: homogeneous stacks ‘train stack’ and ‘rT-coach stack’, heterogeneous stacks ‘atrain stack’ and ‘rA-coach stack’, homogeneous queues ‘train queue’ and ‘rT-coach queue’, heterogeneous queues ‘atrain queue’ and ‘rA-coach queue’, homogeneous binary trees ‘train binary tree’ and ‘rT-coach binary tree’, heterogeneous binary trees ‘atrain binary tree’ and ‘rA-coach binary tree’, homogeneous trees ‘train tree’ and ‘rT-coach tree’, heterogeneous trees ‘atrain tree’ and ‘rA-coach tree’, to enrich the subject ‘Data Structures for Big Data’ for big data science.

INTRODUCTION

The present world of big data are expanding very fast in 4Vs: Volume, Variety, Velocity and Veracity, and also in many more directions. How to deal with explosive momentum of big data, how to process big data in an efficient way within limited resources, etc. are of major concern to the computer scientists now-a-days. For a detailed study about big data, one could see (Berman, 2013; Feinleib, 2013; Needham, 2013; Viktor Mayer-Schönberger, 2013). Big data has variable mass of 4Vs in various directions, and in this sense it is to be philosophically regarded as a vector, or rather as a tensor. The existing data structures of computer science are neither appropriate nor sufficient to deal with the big data. Big data in most of the cases are of heterogeneous type. The homogeneous data structure ‘r-Train’ and the heterogeneous data structure ‘r-Atrain’ for big data are the first attempt to introduce any exclusive data structures (Biswas, 2011, 2012, 2013b, 2015a) for big data.

DOI: 10.4018/978-1-5225-0182-4.ch002

It is obvious that the ‘Data Structures for Big Data’ is to be regarded as a new subject in big data science, not just as a new topic, considering the explosive momentum of the big data in a new universe. One could view big data with philosophical eyes as a tensor of increasing order. The data structures ‘train’ and ‘atrain’ on the platform of infinitely scalable architecture of a new type of distributed system ADS (Biswas, 2015a) designed exclusively for big data, and the new mathematical models: ‘solid matrix/latrix’, ‘solid hematrix/helatrix’, to store big data of any big amount of any datatype have opened easy gates to the world giant organizations and their developers to deal with big data. In this chapter we introduce few very useful data structures for big data which are : homogeneous stacks ‘train stack’ and ‘rT-coach stack’; heterogeneous stacks ‘atrain stack’ and ‘rA-coach stack’; homogeneous queues ‘train queue’ and ‘rT-coach queue’; heterogeneous queues ‘atrain queue’ and ‘rA-coach queue’; homogeneous binary trees ‘train binary tree’ and ‘rT-coach binary tree’; heterogeneous binary trees ‘atrain binary tree’ and ‘rA-coach binary tree’; homogeneous trees ‘train tree’ and ‘rT-coach tree’; heterogeneous trees ‘atrain tree’ and ‘rA-coach tree’; to enrich the subject ‘Data Structures for Big Data’ with the extended notion of the architecture of the data structures r-atrain and r-train for big data. All these data structures, homogeneous or heterogeneous, are designed in compatible with the architecture of a new distributed system ADS (Biswas, 2015a) proposed exclusively for big data of any 4V and infinitely scalable.

PRELIMINARIES OF THE DATA STRUCTURE r-ATRRAIN FOR BIG DATA

The data structure ‘r-Train’ (‘train’, in short) where r is a natural number is a new kind of powerful robust data structure which can store homogeneous big data dynamically in a flexible way. However, the heterogeneous data structure ‘r-Atrain’ (‘atrain’, in short) where r is a natural number is a new kind of powerful robust data structure which can store heterogeneous big data of any 4V dynamically in a flexible way. In (Biswas, 2015a) a detailed study of the two fundamental big data structures train and atrain is available with the detailed description of the distributed system ADS for big data, an infinitely scalable architecture to deal with big data of any 4Vs. However, for the sake of preliminaries of this chapter, the heterogeneous data structure r-atrain for big data is reproduced in brief from (Biswas, 2015a), before we start for the actual content here. For details of train and atrain, one could see (Biswas, 2015a).

The heterogeneous big data structure ‘atrain’ can store heterogeneous data (by the term heterogeneous data, we mean here data of various datatypes). The classical data structures viz. array, linked list, etc. can store and handle with homogeneous data only, not heterogeneous data. The data structure ‘r-Atrain’ (‘atrain’, in short) is one of the most useful, most powerful and most appropriate data structures in the field of Computer Science considering the application potential of it by the programmers programming with any computer languages. This advanced data structure ‘r-Atrain’ (‘atrain’, in short) is logically almost analogous to the homogeneous data structure r-train (train) but with an advanced level of construction to accommodate any variety of heterogeneous big data. The data structure train can be viewed as a special case of the data structure atrain. It is important to note that none of these two new big data structures train and atrain is a competitor of the other, but they are the basic and most fundamental data structures for big data. By default, any heterogeneous data structure can work as a homogeneous data structure too. For working with homogeneous big data, train is more suitable than atrain. However, for working with heterogeneous big data, atrain is suitable while train can not be applicable even. The natural number r is to be suitably predecided and fixed by the programmer depending upon the problem under consideration and also upon the organization/industry for which the big data problem is posed.

26 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/chapter/introducing-data-structures-for-big-data/157682

Related Content

A Temporal Multidimensional Model and OLAP Operators

Waqas Ahmed, Esteban Zimányi, Alejandro Ariel Vaisman and Robert Wrembel (2020). *International Journal of Data Warehousing and Mining* (pp. 112-143).

www.irma-international.org/article/a-temporal-multidimensional-model-and-olap-operators/265260

Efficient and Effective Aggregate Keyword Search on Relational Databases

Luping Li, Stephen Petschulat, Guanting Tang, Jian Pei and Wo-Shun Luk (2012). *International Journal of Data Warehousing and Mining* (pp. 41-81).

www.irma-international.org/article/efficient-effective-aggregate-keyword-search/74755

Data Mining in Decision Support for Bioenergy Production

Nasser Ayoub and Yuji Naka (2009). *Social and Political Implications of Data Mining: Knowledge Management in E-Government* (pp. 235-253).

www.irma-international.org/chapter/data-mining-decision-support-bioenergy/29074

Data Field for Hierarchical Clustering

Shuliang Wang, Wenyan Gan, Deyi Li and Deren Li (2011). *International Journal of Data Warehousing and Mining* (pp. 43-63).

www.irma-international.org/article/data-field-hierarchical-clustering/58637

The Application of Data Mining Techniques in Health Plan Population Management: A Disease Management Approach

Theodore L. Perry, Travis Tucker, Laurel R. Hudson, William Gandy, Amy L. Neftzger and Guy B. Hamar (2004). *Managing Data Mining: Advice from Experts* (pp. 146-166).

www.irma-international.org/chapter/application-data-mining-techniques-health/24783