

Chapter 8

MapReduce and YARN API

ABSTRACT

Apache Hadoop includes Java APIs for different functions on a HDFS file system like creation of a file, renaming, deletion, and to set read-write permissions for directories. This can be done on a single and cluster of systems. In addition, REST (REpresentational State Transfer) APIs is a collection of web services to provide interoperability between a single system and an interconnected distributed network. REST is chosen for its speedy performance, scalability, simplicity, and reliability. YARN REST and MapReduce REST APIs are briefly discussed in this chapter. YARN web service REST API includes URI resources through which the cluster information, nodes, and application information can be accessed. YARN is comprised of Resource manager, node manager, and timeline REST APIs. The application has HTTP request as resource and the response can be in the form XML or JSON. The request URI, response status, header, and body are defined in actual format. Similarly, the REST API is used for MapReduce that comprises the details about the jobs running with the information such as number of tasks, counters, and attempts. Hence, the REST APIs on YARN and resource manager create small modules as a response when a resource is requested. An outline of the research and growth of REST APIs is included in this chapter.

INTRODUCTION

Hadoop provides many Java native APIs that support file system operations like creation of a file or renaming or deletion and operations with directories

DOI: 10.4018/978-1-5225-3790-8.ch008

to set read, write permissions etc... Hadoop cluster requires these operations to manipulate the data in files across HDFS. This is supported by an additional API based on REST functionalities to map these requirements. The Hadoop YARN and Mapreduce REST API's support various services across applications working in the environment.

BACKGROUND

REpresentational State Transfer (REST) is web service to provide interoperability between a single system and an interconnected distributed network. It allows the requesting system to access and manipulate data of web resources using uniform set of stateless operations. In a REST API, requests will be in the form of resource URI which may elicit a response in XML, JSON, and HTML etc... REST is chosen for its fast performance, scalability, visibility, simplicity, reliability, reusing components and gets the system updated without affecting it. The services which adhere to the architectural constraints and properties of REST APIs utilize it. One such case is that Hadoop services that combine the architecture of YARN and Mapreduce based RESTful APIs.

REST uses HTTP protocol for communication in the web world. URI is used for communication among the resources of RESTful services. The HTTP methods supported are:

1. **GET:** Read a resource (Read only).
2. **PUT:** Create new resource.
3. **POST:** Update an existing resource or create a new resource.
4. **DELETE:** Remove a resource.
5. **OPTIONS:** Get the supported operations on the resource.

YARN REST APIs

Hadoop YARN web service REST APIs includes set of URI resources through which the cluster information, nodes and application information can be accessed. The resources can be grouped based on type of information. Some can be together while others using URI resource

The URI of REST based web service is:

20 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/mapreduce-and-yarn-api/216603

Related Content

A Survey on Grey Optimization

Adem Guluma Negewo (2018). *Optimization Techniques for Problem Solving in Uncertainty* (pp. 1-30).

www.irma-international.org/chapter/a-survey-on-grey-optimization/206628

On Estimating Population Means of Two-Sensitive Variables With Ranked Set Sampling Design

Agustin Santiago Moreno and Khalid Ul Islam Rather (2022). *Ranked Set Sampling Models and Methods* (pp. 42-61).

www.irma-international.org/chapter/on-estimating-population-means-of-two-sensitive-variables-with-ranked-set-sampling-design/291278

Comparison of Graduates' and Academics' Perceptions of the Skills Required for Big Data Analysis: Statistics Education in the Age of Big Data

Busayasachee Puang-Ngern, Ayse A. Bilgin and Timothy J. Kyng (2017). *Data Visualization and Statistical Literacy for Open and Big Data* (pp. 126-152).

www.irma-international.org/chapter/comparison-of-graduates-and-academics-perceptions-of-the-skills-required-for-big-data-analysis/179964

Big Data, 3D Printing Technology, and Industry of the Future

Micheal Omotayo Alabi (2017). *International Journal of Big Data and Analytics in Healthcare* (pp. 1-20).

www.irma-international.org/article/big-data-3d-printing-technology-and-industry-of-the-future/204445

Big Data Impacts in Tourism Research

Arjun Singh, Surbhi Chauhan, Nitin Sharma, Vijay Kumar Sharma and Raguru Jaya Krishna (2021). *Big Data Analytics for Improved Accuracy, Efficiency, and Decision Making in Digital Marketing* (pp. 74-84).

www.irma-international.org/chapter/big-data-impacts-in-tourism-research/280644