

Chapter 6

YARN

ABSTRACT

Apache Hadoop YARN (Yet Another Resource Negotiator) is the cluster resource management technology in Hadoop version 2. The YARN provides multiple accesses for batch and real-time processing across multiple clusters and has the benefit over utilization of cluster resources during dynamic allocation. The chapter shows the YARN architecture, schedulers, resource manager phases, YARN applications, commands, and timeline server. The architecture of YARN splits the task into resource management and job scheduling. This is managed by Resource Manager and Node Manager. The chapter addresses the Timeline Server, which stores and retrieves the present and past information in a generic way.

INTRODUCTION

YARN as an architectural core of Hadoop is a platform for consistent operations that allows data processing operations like streaming, real time analytics and governance across multiple clusters. Hadoop cluster feature multi-tenancy is enhanced by YARN improving the enterprise's return with multiple accesses for batch, interactive and real time engines working on a same data at a particular time. YARN perk up the cluster utilization with dynamic allocation of cluster resources. Data center expands for processing to thousands of nodes handling zettabytes of data. YARN's resource manager centres wholly on scheduling and highly compatible than mapreduce used in Hadoop 1.

DOI: 10.4018/978-1-5225-3790-8.ch006

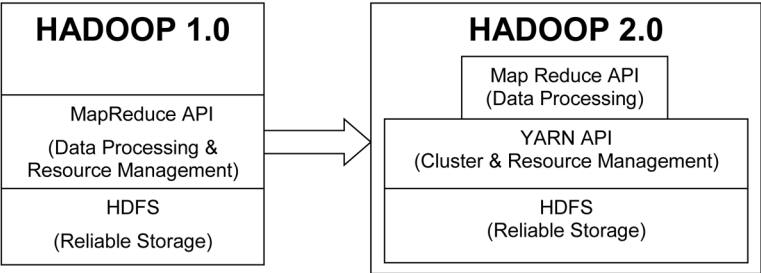
BACKGROUND

Hadoop being the foundation of the big data era, there is processing model difference between hadoop 1 and hadoop 2. Hadoop 1 includes the progress with HDFS for storage and the processing by batch oriented MapReduce jobs. The version 1 is potential for distributed processing but not much suitable for interactive analysis along with memory intensive algorithms. Hence hadoop 2 includes new version of HDFS federation and resource manager YARN. HDFS federation combines the measure of scalability and reliability to hadoop, YARN supports and implements a flexible execution engine with high end processing models. It further separates processing and resource management of Mapreduce in Hadoop 1. Additionally, it is responsible for administering workloads with security controls and to maintain multi tenant environment amid high availability features.

YARN ARCHITECTURE

YARN (Yet Another Resource Negotiator) is split up with two major tasks i.e. resource management and job scheduling to act on a global environment and as per application. An application can either refer to a job or a DAG of jobs. In Hadoop 1, inflexible slots are configured on nodes which gets underutilized when more map or reduce tasks are running and also can't share resources with non map reduce applications running on Hadoop cluster like Impala, Giraph etc. A host in Hadoop refers to a node and cluster is the connection of two or more nodes joined by a high speed network. Nodes may be partitioned in racks and blocks as in hardware part of hadoop infrastructure. There can be thousands of hosts in a cluster. In Hadoop, there is one master

Figure 1. Hadoop 1 => Hadoop 2 Comparison



33 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/yarn/216601

Related Content

Census Data Analysis and Visualization Using R Tool: A Case Study

Veena Gadadand Sowmyarani C. N. (2019). *Managerial Perspectives on Intelligent Big Data Analytics* (pp. 137-161).

www.irma-international.org/chapter/census-data-analysis-and-visualization-using-r-tool/224336

Descriptive Data Analytics on Dinesafe Data for Food Assessment and Evaluation Using R Programming Language: A Case Study on Toronto's Dinesafe Inspection and Disclosure System

Ajinkya Kunjir, Jugal Shahand Vikas Trikha (2021). *Handbook of Research on Engineering, Business, and Healthcare Applications of Data Science and Analytics* (pp. 485-507).

www.irma-international.org/chapter/descriptive-data-analytics-on-dinesafe-data-for-food-assessment-and-evaluation-using-r-programming-language/264325

Big Data Applications in Vaccinology

Joseph E. Kasten (2021). *International Journal of Big Data and Analytics in Healthcare* (pp. 59-80).

www.irma-international.org/article/big-data-applications-in-vaccinology/276927

Agent-Based Modelling in Multicellular Systems Biology

Sara Montagnaand Andrea Omicini (2020). *Data Analytics in Medicine: Concepts, Methodologies, Tools, and Applications* (pp. 369-389).

www.irma-international.org/chapter/agent-based-modelling-in-multicellular-systems-biology/243121

A High-Level Interactive Query Language for Big Data Analytics Based on a Functional Model

Symphorien Monsiaand Sami Faiz (2020). *International Journal of Data Analytics* (pp. 22-37).

www.irma-international.org/article/a-high-level-interactive-query-language-for-big-data-analytics-based-on-a-functional-model/244167