Apache Hadoop YARN: Yet Another Resource Negotiator

Hadoop clusters are the common place where engineers and researchers have instantaneous and near to unrestricted access to huge bunch of data which are sorted and processed. For the Hadoop's success, it works as a cause as well as curse. Hence, to tackle this problem and eliminate the misuses and limitations, author represents the next generation of Hadoop's compute Platform, which is known as YARN.

This paper describes the design, development, and current state of deployment of it. It delegates many scheduling related functions to per job components by separation of programming model and resource management functions. In this paper author mentioned some requirements which are needed in YARN such as scalability, multi-tenancy, serviceability, Locality awareness, high cluster utilization, Reliability or Availability, secure and auditable operation, support for programming model diversity, flexible resource model, and backward compatibility. To fulfill these requirements YARN lifts some functions into a platform layer. ResourceManager is responsible for the resource management and ApplicationMaster is responsible for leaving coordination of logical execution plans to a host of framework implementations. YARN comes with numerous improvements such as greater scalability, higher efficiency and to efficiently share a cluster it allows enormous number of different frameworks. In this paper, author provides experimental evidences which are demonstrating these improvements. In the real world, these all advantages are in a both ways experimentally as well as practically claimed by Yahoo!, which is now fully running on YARN. One of the main aspects of YARN is that, to manage containers and communications it allows frameworks built on top of it. It has a request-based approach and based on various criteria it allows the ApplicationMaster (AM) to ask for resources. Based on what was given and current usage-it allows requester to modify future requests. According to the all data, we can conclude that due to separation YARN is useful in both direction, huge production framework as well as it can be served as an invaluable Platform for the research community.