Ethics Pledge

Consistent with the above statements, all homework exercises, tests and exams that are designated as individual assignments MUST contain the following signed statement before they can be accepted for grading.

I pledge on my honor that I have not given or received any unauthorized assistance on this assignment/examination. I further pledge that I have not copied any material from a book, article, the Internet or any other source except where I have expressly cited the source.

Signature: Haodong Zhao Date: Feb 12th. 2019

Please note that assignments in this class may be submitted to

www.turnitin.com, a web-based anti-plagiarism system, for an evaluation of their originality.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Reading review**

**Apache Hadoop YARN: Yet Another Resource Negotiator**

At present, the widespread use of Hadoop has exceeded its expectations, thus revealing 2 key shortcomings:

1. The specific programming model is tightly coupled to the resource management infrastructure, allowing developers to abuse the MapReduce programming model.
2. Centralize the control flow of the job, allowing the scheduler to expand endlessly.

This article introduces the next generation Hadoop computing platform: YARN’s design, development and current state.

Apache Hadoop was originally one of many open source implementations of MapReduce, focusing on the unprecedented scale required to index web crawlers. Focus on the powerful fault tolerance of large-scale data-intensive computing. And it allows engineers and researchers to ask a lot of computing resources and a lot of company data almost without restrictions.

As a next-generation Hadoop computing platform, YARN is separated from the familiar monolithic architecture. By separating the resource management functions from the programming model, YARN delegates a lot of scheduling-related functions to each job, providing great flexibility.

Then, the article introduces the history and principles of Apache Hadoop. From the era of ad-hoc clusters to MapReduce cluster, and then to the YARN method.

Based on the lessons of the evolving Apache Hadoop MapReduce, YARN is designed to meet the following requirements:

1. Scalability
2. Revolving around multi-tenant needs
3. Maintainability
4. The aspect of local consciousness
5. High cluster utilization
6. Group problems based on reliability, availability, and needs
7. Security and auditable operations must be maintained in YARN
8. YARN must declare a truce to its users and provide clear support for the diversity of programming models
9. The need for a flexible resource model
10. Backward compatibility

Finally, the article describes what kind of architecture should be implemented in order to achieve the above requirements of YARN. And the application of YARN in the real world and several popular frameworks that have been ported to YARN.

Through the recall of Hadoop history and its application, the author describes to us the gradual but profound transformation leading to YARN. And believes that YARN can be used as reliable production framework or as a valuable experience in the research community.