Wyatt Blair

EMT-678-WS: Big Data Technologies

10/3/2024

**Class 3—Apache Hadoop YARN: Yet Another Resource Negotiator**

Apache Hadoop YARN is a methodology in which system resource management and job management is separated and linked using clever MapReduce jobs. In general, there is a Resource Manager (RM), which handles the task of resource management, and an Application Manager (AM), which handles job management. ­­­The paper lists ten requirements for Yahoo! In its adoption of Apache Hadoop:

1. Scalability
2. Multi-Tenancy
3. Serviceability
4. Locality Awareness
5. High Cluster Utilization
6. Reliability / Availability
7. Secure and Auditable Operation
8. Support for Programming Model Diversity
9. Flexible Resource Model
10. Backward Compatibility

The paper goes into a good amount of detail describing how and why each of these ten requirements came to be and what problems they solve specifically. YARN is the solution to all of them. The gist of how YARN works is as follows: the per-cluster RM tracks global resource usage and ensures the whole framework is operating smoothly within the physical constraints of the system. In tandem, the AM creates a logical plan for a single job by requesting resources from the RM, generates a physical plan once it learns what resources are available, and coordinates the execution that plan around any potential faults. This structure, facilitated by the MapReduce algorithm, allows for all ten requirements to be met in an efficient manner.