Scheduling Algorithms in MapReduce

Distributed Mind

Karthik Kotian, Jason A Smith, Ye Zhang
Overview of topic

MapReduce
- Process a large set of data in parallel across a distributed cluster of computers
- Maximize the speed at which a task is completed

Scheduling
- Splits a task into smaller jobs and distributes work to the nodes in the cluster
- Utilize the cluster's resources to its maximum capacity
MapReduce

Schedulers

Algorithms

- **FIFO**\(^2\) - Oldest Job First.
- **Fairness**\(^2\) - Every Task gets equal share of resources.
- **Capacity**\(^2\) - Built over Fairness for Large Cluster.
  - Queues
  - Priorities
- **Adaptive**\(^1\) - Balance between utilization and parallelism in the cluster.
- **Data locality**\(^3\) - Minimize data movement.

1- Research Paper #1, #2 - Research Paper #2, #3 - Research Paper #3
Hypothesis

- **FIFO**
  - Best when \# of tasks < cluster size

- **Fairness**
  - Best when cluster size < \# of tasks

- **Capacity**
  - Best when there's multiple tasks and priorities as measured by average response time
Approach to test hypothesis

- Simulate MapReduce with each scheduler.
- Change the:
  - Size of the cluster.
  - Complexity of the tasks.
  - Number of tasks.
  - Priority of the tasks.
- Compare schedulers based on average task completion time.
Software to deliver

- Java SE 6
- **Application:**
  - Master Node object
  - Worker Node object
  - "Task" object
    - Complexity (how long it takes to process)
    - File/Data Size
    - Priority
    - Split into Jobs
  - Scheduler Object
    - Queue
Aysan Rasooli, Douglas G. Down

An Adaptive Scheduling Algorithm for Dynamic Heterogeneous Hadoop Systems

Proceedings of the 2011 Conference of the Center for Advanced Studies on Collaborative Research

2011

Pages 30-44

http://www.cas.mcmaster.ca/~downd/cascon11.pdf
M. Tim Jones

**Scheduling in Hadoop** - An introduction to the pluggable scheduler framework

IBM Technical Library

06 Dec 2011

Pages 1 -13

Zhenhua Guo, Geoffrey, Mo Zhou, Indiana University, Bloomington, IL

Investigation of Data Locality and Fairness in MapReduce

MapReduce 2012 Proceeding of third international workshop on MapReduce and its application date.

2012

Page 25 - 32

Questions