YARN Architecture

Tuesday, January 10, 2023   7:10 AM

- Architecture
  - Use Case / Motivation
    - Cluster resource management / scheduling system
    - Framework for job submission / reporting / control & service / Daemon
    - Multi tenancy

    - Utilization & Efficient Allocation of Compute (Time / Task)
    - Minimize I/O Cost & Transit when not needed (Overhead)

    - Security, Maintenance, etc...

  - Resource Management
    - CPU / GPU (Extensions)
    - Memory
    - Disk
    - Network } (Extensions)

  * Constraints
    - Node Locality
    - Rack Locality
    - Tags (Generic)

      ↳ Time Constraints: End Time, Run Time, Start Time (Fairness)

  - Components
  - Cluster: n ≥ 2 nodes, LAN / WAN
    - Master Nodes   Control / Client Connectivity
    - Worker Nodes   Tasks (vmem, vcores)

    - Manager Instances
      - Resource Manager   per Cluster
      - Node Manager   per Node

    - Containers
      - Requested Resource Requirements
      - Allocation: Find nodes / hosts to "hold" a container!

    - Application
      - Client: Driver (How to run)
      - Application Master: Allows for resource usage on YARN nodes
      - Tasks: Map, Reduce